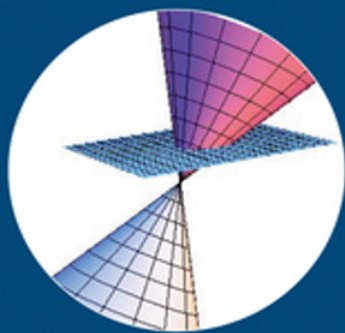
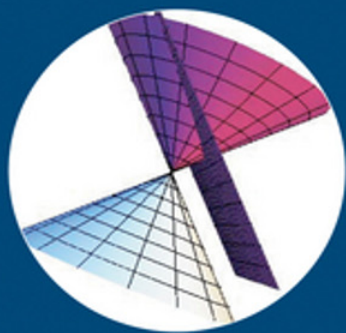
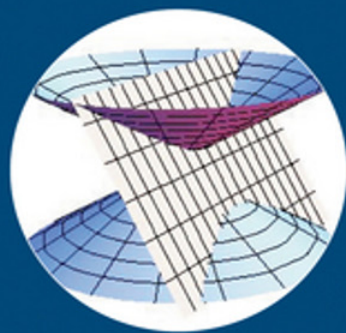


THIRD EDITION

THE HISTORY OF MATHEMATICS

A BRIEF COURSE



ROGER L. COOKE

 WILEY

THE HISTORY OF MATHEMATICS

THE HISTORY OF MATHEMATICS A BRIEF COURSE

THIRD EDITION

Roger L. Cooke

Department of Mathematics and Statistics
University of Vermont
Burlington, VT



A JOHN WILEY & SONS, INC., PUBLICATION

Copyright © 2013 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

Cooke, Roger, 1942-

The history of mathematics : a brief course / Roger L. Cooke. – 3rd ed.

p. cm.

Includes bibliographical references and index.

ISBN 978-1-118-21756-6 (cloth)

1. Mathematics–History. I. Title.

QA21.C649 2013

510'.9–dc23

2012020963

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

CONTENTS

PREFACE	xxiii
Changes from the Second Edition	xxiii
Elementary Texts on the History of Mathematics	xxiv
PART I. WHAT IS MATHEMATICS?	
Contents of Part I	1
1. Mathematics and its History	3
1.1. Two Ways to Look at the History of Mathematics	3
1.1.1. History, but not Heritage	4
1.1.2. Our Mathematical Heritage	4
1.2. The Origin of Mathematics	5
1.2.1. Number	5
1.2.2. Space	5
1.2.3. Are Mathematical Ideas Innate?	7
1.2.4. Symbolic Notation	7
1.2.5. Logical Relations	7
1.2.6. The Components of Mathematics	8
1.3. The Philosophy of Mathematics	8
1.3.1. Mathematical Analysis of a Real-World Problem	9
1.4. Our Approach to the History of Mathematics	11
Questions for Reflection	12
2. Proto-mathematics	14
2.1. Number	14
2.1.1. Animals' Use of Numbers	14
2.1.2. Young Children's Use of Numbers	15
2.1.3. Archaeological Evidence of Counting	15
2.2. Shape	16
2.2.1. Perception of Shape by Animals	16
2.2.2. Children's Concepts of Space	16
2.2.3. Geometry in Arts and Crafts	17
2.3. Symbols	18

2.4.	Mathematical Reasoning	20
2.4.1.	Animal Reasoning	20
2.4.2.	Visual Reasoning	21
	Problems and Questions	22
	Mathematical Problems	22
	Questions for Reflection	24
PART II. THE MIDDLE EAST, 2000–1500 BCE		
	Contents of Part II	25
3.	Overview of Mesopotamian Mathematics	27
3.1.	A Sketch of Two Millennia of Mesopotamian History	27
3.2.	Mathematical Cuneiform Tablets	29
3.3.	Systems of Measuring and Counting	30
3.3.1.	Counting	31
3.4.	The Mesopotamian Numbering System	31
3.4.1.	Place-Value Systems	32
3.4.2.	The Sexagesimal Place-Value System	33
3.4.3.	Converting a Decimal Number to Sexagesimal	33
3.4.4.	Irrational Square Roots	36
	Problems and Questions	36
	Mathematical Problems	36
	Historical Questions	36
	Questions for Reflection	37
4.	Computations in Ancient Mesopotamia	38
4.1.	Arithmetic	38
4.1.1.	Square Roots	39
4.2.	Algebra	40
4.2.1.	Linear and Quadratic Problems	41
4.2.2.	Higher-Degree Problems	43
	Problems and Questions	44
	Mathematical Problems	44
	Historical Questions	44
	Questions for Reflection	44
5.	Geometry in Mesopotamia	46
5.1.	The Pythagorean Theorem	46
5.2.	Plane Figures	48
5.2.1.	Mesopotamian Astronomy	48
5.3.	Volumes	49
5.4.	Plimpton 322	49
5.4.1.	The Purpose of Plimpton 322: Some Conjectures	53

Problems and Questions	54
Mathematical Problems	54
Historical Questions	55
Questions for Reflection	55
6. Egyptian Numerals and Arithmetic	56
6.1. Sources	56
6.1.1. Mathematics in Hieroglyphics and Hieratic	57
6.2. The Rhind Papyrus	58
6.3. Egyptian Arithmetic	58
6.4. Computation	59
6.4.1. Multiplication and Division	61
6.4.2. “Parts”	62
Problems and Questions	65
Mathematical Problems	65
Historical Questions	65
Questions for Reflection	65
7. Algebra and Geometry in Ancient Egypt	66
7.1. Algebra Problems in the Rhind Papyrus	66
7.1.1. Applied Problems: The <i>Pesu</i>	67
7.2. Geometry	68
7.3. Areas	69
7.3.1. Rectangles, Triangles, and Trapezoids	69
7.3.2. Slopes	69
7.3.3. Circles	70
7.3.4. The Pythagorean Theorem	71
7.3.5. Spheres or Cylinders?	72
7.3.6. Volumes	73
Problems and Questions	76
Mathematical Problems	76
Historical Questions	76
Questions for Reflection	76
PART III. GREEK MATHEMATICS FROM 500 BCE TO 500 CE	
Contents of Part III	77
8. An Overview of Ancient Greek Mathematics	79
8.1. Sources	80
8.1.1. Loss and Recovery	81
8.2. General Features of Greek Mathematics	82
8.2.1. Pythagoras	83
8.2.2. Mathematical Aspects of Plato’s Philosophy	85

8.3.	Works and Authors	87
8.3.1.	Euclid	87
8.3.2.	Archimedes	87
8.3.3.	Apollonius	88
8.3.4.	Zenodorus	88
8.3.5.	Heron	88
8.3.6.	Ptolemy	89
8.3.7.	Diophantus	89
8.3.8.	Pappus	89
8.3.9.	Theon and Hypatia	89
	Questions	90
	Historical Questions	90
	Questions for Reflection	90
9.	Greek Number Theory	91
9.1.	The Euclidean Algorithm	92
9.2.	The <i>Arithmetica</i> of Nicomachus	93
9.2.1.	Factors vs. Parts. Perfect Numbers	94
9.2.2.	Figurate Numbers	95
9.3.	Euclid's Number Theory	97
9.4.	The <i>Arithmetica</i> of Diophantus	97
9.4.1.	Algebraic Symbolism	98
9.4.2.	Contents of the <i>Arithmetica</i>	99
9.4.3.	Fermat's Last Theorem	100
	Problems and Questions	101
	Mathematical Problems	101
	Historical Questions	102
	Questions for Reflection	102
10.	Fifth-Century Greek Geometry	103
10.1.	"Pythagorean" Geometry	103
10.1.1.	Transformation and Application of Areas	103
10.2.	Challenge No. 1: Unsolved Problems	106
10.3.	Challenge No. 2: The Paradoxes of Zeno of Elea	107
10.4.	Challenge No. 3: Irrational Numbers and Incommensurable Lines	108
10.4.1.	The Arithmetical Origin of Irrationals	110
10.4.2.	The Geometric Origin of Irrationals	110
10.4.3.	Consequences of the Discovery	111
	Problems and Questions	113
	Mathematical Problems	113
	Historical Questions	113
	Questions for Reflection	114

11. Athenian Mathematics I: The Classical Problems	115
11.1. Squaring the Circle	116
11.2. Doubling the Cube	117
11.3. Trisecting the Angle	122
11.3.1. A Mechanical Solution: The Conchoid	125
Problems and Questions	126
Mathematical Problems	126
Historical Questions	126
Questions for Reflection	127
12. Athenian Mathematics II: Plato and Aristotle	128
12.1. The Influence of Plato	128
12.2. Eudoxan Geometry	130
12.2.1. The Eudoxan Definition of Proportion	130
12.2.2. The Method of Exhaustion	131
12.2.3. Ratios in Greek Geometry	133
12.3. Aristotle	134
Problems and Questions	138
Mathematical Problems	138
Historical Questions	138
Questions for Reflection	139
13. Euclid of Alexandria	140
13.1. The <i>Elements</i>	140
13.1.1. Book 1	141
13.1.2. Book 2	141
13.1.3. Books 3 and 4	143
13.1.4. Books 5 and 6	143
13.1.5. Books 7–9	143
13.1.6. Book 10	144
13.1.7. Books 11–13	144
13.2. The <i>Data</i>	144
Problems and Questions	145
Mathematical Problems	145
Historical Questions	147
Questions for Reflection	147
14. Archimedes of Syracuse	148
14.1. The Works of Archimedes	149
14.2. The Surface of a Sphere	150
14.3. The Archimedes Palimpsest	153
14.3.1. The <i>Method</i>	154

14.4. Quadrature of the Parabola	155
14.4.1. The Mechanical Quadrature	155
14.4.2. The Rigorous Quadrature	156
Problems and Questions	158
Mathematical Problems	158
Historical Questions	158
Questions for Reflection	159
15. Apollonius of Perga	160
15.1. History of the <i>Conics</i>	161
15.2. Contents of the <i>Conics</i>	162
15.2.1. Properties of the Conic Sections	165
15.3. Foci and the Three- and Four-Line Locus	165
Problems and Questions	166
Mathematical Problems	166
Historical Questions	168
Questions for Reflection	168
16. Hellenistic and Roman Geometry	169
16.1. Zenodorus	169
16.2. The Parallel Postulate	171
16.3. Heron	172
16.4. Roman Civil Engineering	174
Problems and Questions	176
Mathematical Problems	176
Historical Questions	176
Questions for Reflection	176
17. Ptolemy's Geography and Astronomy	177
17.1. Geography	177
17.2. Astronomy	180
17.2.1. Epicycles and Eccentrics	181
17.2.2. The Motion of the Sun	182
17.3. The <i>Almagest</i>	184
17.3.1. Trigonometry	184
17.3.2. Ptolemy's Table of Chords	184
Problems and Questions	187
Mathematical Problems	187
Historical Questions	188
Questions for Reflection	188
18. Pappus and the Later Commentators	190
18.1. The <i>Collection</i> of Pappus	190
18.1.1. Generalization of the Pythagorean Theorem	191

18.1.2. The Isoperimetric Problem	191
18.1.3. Analysis, Locus Problems, and Pappus' Theorem	191
18.2. The Later Commentators: Theon and Hypatia	196
18.2.1. Theon of Alexandria	196
18.2.2. Hypatia of Alexandria	197
Problems and Questions	198
Mathematical Problems	198
Historical Questions	199
Questions for Reflection	199
PART IV. INDIA, CHINA, AND JAPAN 500 BCE–1700 CE	
Contents of Part IV	201
19. Overview of Mathematics in India	203
19.1. The <i>Sulva Sutras</i>	205
19.2. Buddhist and Jain Mathematics	206
19.3. The Bakshali Manuscript	206
19.4. The <i>Siddhantas</i>	206
19.5. Hindu–Arabic Numerals	206
19.6. Aryabhata I	207
19.7. Brahmagupta	208
19.8. Bhaskara II	209
19.9. Muslim India	210
19.10. Indian Mathematics in the Colonial Period and After	210
19.10.1. Srinivasa Ramanujan	210
Questions	211
Historical Questions	211
Questions for Reflection	211
20. From the <i>Vedas</i> to Aryabhata I	213
20.1. Problems from the <i>Sulva Sutras</i>	213
20.1.1. Arithmetic	213
20.1.2. Geometry	214
20.1.3. Square Roots	216
20.1.4. Jain Mathematics: The Infinite	217
20.1.5. Jain Mathematics: Combinatorics	217
20.1.6. The Bakshali Manuscript	218
20.2. Aryabhata I: Geometry and Trigonometry	219
20.2.1. Trigonometry	220
20.2.2. The <i>Kuttaka</i>	224
Problems and Questions	225
Mathematical Problems	225
Historical Questions	225
Questions for Reflection	225

21. Brahmagupta, the <i>Kuttaka</i>, and Bhaskara II	227
21.1. Brahmagupta's Plane and Solid Geometry	227
21.2. Brahmagupta's Number Theory and Algebra	228
21.2.1. Pythagorean Triples	229
21.2.2. Pell's Equation	229
21.3. The <i>Kuttaka</i>	230
21.4. Algebra in the Works of Bhaskara II	233
21.4.1. The <i>Vija Ganita (Algebra)</i>	233
21.4.2. Combinatorics	233
21.5. Geometry in the Works of Bhaskara II	235
Problems and Questions	237
Mathematical Problems	237
Historical Questions	238
Questions for Reflection	238
22. Early Classics of Chinese Mathematics	239
22.1. Works and Authors	240
22.1.1. The <i>Zhou Bi Suan Jing</i>	241
22.1.2. The <i>Jiu Zhang Suan Shu</i>	242
22.1.3. The <i>Sun Zi Suan Jing</i>	242
22.1.4. Liu Hui. The <i>Hai Dao Suan Jing</i>	242
22.1.5. Zu Chongzhi and Zu Geng	243
22.1.6. Yang Hui	243
22.1.7. Cheng Dawei	243
22.2. China's Encounter with Western Mathematics	243
22.3. The Chinese Number System	244
22.3.1. Fractions and Roots	245
22.4. Algebra	246
22.5. Contents of the <i>Jiu Zhang Suan Shu</i>	247
22.6. Early Chinese Geometry	249
22.6.1. The <i>Zhou Bi Suan Jing</i>	249
22.6.2. The <i>Jiu Zhang Suan Shu</i>	251
22.6.3. The <i>Sun Zi Suan Jing</i>	253
Problems and Questions	253
Mathematical Problems	253
Historical Questions	253
Questions for Reflection	253
23. Later Chinese Algebra and Geometry	255
23.1. Algebra	255
23.1.1. Systems of Linear Equations	256
23.1.2. Quadratic Equations	256
23.1.3. Cubic Equations	257
23.1.4. A Digression on the Numerical Solution of Equations	258

23.2. Later Chinese Geometry	262
23.2.1. Liu Hui	262
23.2.2. Zu Chongzhi	264
Problems and Questions	265
Mathematical Problems	265
Historical Questions	266
Questions for Reflection	266
24. Traditional Japanese Mathematics	267
24.1. Chinese Influence and Calculating Devices	267
24.2. Japanese Mathematicians and Their Works	268
24.2.1. Yoshida Koyu	269
24.2.2. Seki Kōwa and Takebe Kenkō	269
24.2.3. The Modern Era in Japan	270
24.3. Japanese Geometry and Algebra	270
24.3.1. Determinants	272
24.3.2. The Challenge Problems	273
24.3.3. Beginnings of the Calculus in Japan	274
24.4. <i>Sangaku</i>	277
24.4.1. Analysis	279
Problems and Questions	279
Mathematical Problems	279
Historical Questions	280
Questions for Reflection	280
PART V. ISLAMIC MATHEMATICS, 800–1500	
Contents of Part V	281
25. Overview of Islamic Mathematics	283
25.1. A Brief Sketch of the Islamic Civilization	283
25.1.1. The Umayyads	283
25.1.2. The Abbasids	284
25.1.3. The Turkish and Mongol Conquests	284
25.1.4. The Islamic Influence on Science	284
25.2. Islamic Science in General	285
25.2.1. Hindu and Hellenistic Influences	285
25.3. Some Muslim Mathematicians and Their Works	287
25.3.1. Muhammad ibn Musa al-Khwarizmi	287
25.3.2. Thabit ibn-Qurra	287
25.3.3. Abu Kamil	288
25.3.4. Al-Battani	288
25.3.5. Abu'l Wafa	288
25.3.6. Ibn al-Haytham	288
25.3.7. Al-Biruni	289

25.3.8. Omar Khayyam	289
25.3.9. Sharaf al-Tusi	289
25.3.10. Nasir al-Tusi	289
Questions	290
Historical Questions	290
Questions for Reflection	290
26. Islamic Number Theory and Algebra	292
26.1. Number Theory	292
26.2. Algebra	294
26.2.1. Al-Khwarizmi	295
26.2.2. Abu Kamil	297
26.2.3. Omar Khayyam	297
26.2.4. Sharaf al-Din al-Tusi	299
Problems and Questions	300
Mathematical Problems	300
Historical Questions	301
Questions for Reflection	301
27. Islamic Geometry	302
27.1. The Parallel Postulate	302
27.2. Thabit ibn-Qurra	302
27.3. Al-Biruni: Trigonometry	304
27.4. Al-Kuhi	305
27.5. Al-Haytham and Ibn-Sahl	305
27.6. Omar Khayyam	307
27.7. Nasir al-Din al-Tusi	308
Problems and Questions	309
Mathematical Problems	309
Historical Questions	309
Questions for Reflection	310
PART VI. EUROPEAN MATHEMATICS, 500–1900	
Contents of Part VI	311
28. Medieval and Early Modern Europe	313
28.1. From the Fall of Rome to the Year 1200	313
28.1.1. Boethius and the Quadrivium	313
28.1.2. Arithmetic and Geometry	314
28.1.3. Music and Astronomy	315
28.1.4. The Carolingian Empire	315

28.1.5. Gerbert	315
28.1.6. Early Medieval Geometry	317
28.1.7. The Translators	318
28.2. The High Middle Ages	318
28.2.1. Leonardo of Pisa	319
28.2.2. Jordanus Nemorarius	319
28.2.3. Nicole d'Oresme	319
28.2.4. Regiomontanus	320
28.2.5. Nicolas Chuquet	320
28.2.6. Luca Pacioli	320
28.2.7. Leon Battista Alberti	321
28.3. The Early Modern Period	321
28.3.1. Scipione del Ferro	321
28.3.2. Niccolò Tartaglia	321
28.3.3. Girolamo Cardano	321
28.3.4. Ludovico Ferrari	322
28.3.5. Rafael Bombelli	322
28.4. Northern European Advances	322
28.4.1. François Viète	322
28.4.2. John Napier	322
Questions	323
Historical Questions	323
Questions for Reflection	323
29. European Mathematics: 1200–1500	324
29.1. Leonardo of Pisa (Fibonacci)	324
29.1.1. The <i>Liber abaci</i>	324
29.1.2. The Fibonacci Sequence	325
29.1.3. The <i>Liber quadratorum</i>	326
29.1.4. The <i>Flos</i>	327
29.2. Hindu–Arabic Numerals	328
29.3. Jordanus Nemorarius	329
29.4. Nicole d'Oresme	330
29.5. Trigonometry: Regiomontanus and Pitiscus	331
29.5.1. Regiomontanus	331
29.5.2. Pitiscus	332
29.6. A Mathematical Skill: <i>Prosthaphæresis</i>	333
29.7. Algebra: Pacioli and Chuquet	335
29.7.1. Luca Pacioli	335
29.7.2. Chuquet	335
Problems and Questions	336
Mathematical Problems	336
Historical Questions	337
Questions for Reflection	337

30. Sixteenth-Century Algebra	338
30.1. Solution of Cubic and Quartic Equations	338
30.1.1. Ludovico Ferrari	339
30.2. Consolidation	340
30.2.1. François Viète	341
30.3. Logarithms	343
30.3.1. Arithmetical Implementation of the Geometric Model	344
30.4. Hardware: Slide Rules and Calculating Machines	345
30.4.1. The Slide Rule	345
30.4.2. Calculating Machines	345
Problems and Questions	346
Mathematical Problems	346
Historical Questions	346
Questions for Reflection	346
31. Renaissance Art and Geometry	348
31.1. The Greek Foundations	348
31.2. The Renaissance Artists and Geometers	349
31.3. Projective Properties	350
31.3.1. Girard Desargues	352
31.3.2. Blaise Pascal	355
Problems and Questions	356
Mathematical Problems	356
Historical Questions	357
Questions for Reflection	357
32. The Calculus Before Newton and Leibniz	358
32.1. Analytic Geometry	358
32.1.1. Pierre de Fermat	359
32.1.2. René Descartes	359
32.2. Components of the Calculus	363
32.2.1. Tangent and Maximum Problems	363
32.2.2. Lengths, Areas, and Volumes	365
32.2.3. Bonaventura Cavalieri	365
32.2.4. Gilles Personne de Roberval	366
32.2.5. Rectangular Approximations and the Method of Exhaustion	367
32.2.6. Blaise Pascal	368
32.2.7. The Relation Between Tangents and Areas	370
32.2.8. Infinite Series and Products	370
32.2.9. The Binomial Series	371
Problems and Questions	371
Mathematical Problems	371
Historical Questions	372
Questions for Reflection	372

33. Newton and Leibniz	373
33.1. Isaac Newton	373
33.1.1. Newton's First Version of the Calculus	373
33.1.2. Fluxions and Fluents	374
33.1.3. Later Exposition of the Calculus	374
33.1.4. Objections	375
33.2. Gottfried Wilhelm von Leibniz	375
33.2.1. Leibniz' Presentation of the Calculus	376
33.2.2. Later Reflections on the Calculus	378
33.3. The Disciples of Newton and Leibniz	379
33.4. Philosophical Issues	379
33.4.1. The Debate on the Continent	380
33.5. The Priority Dispute	381
33.6. Early Textbooks on Calculus	382
33.6.1. The State of the Calculus Around 1700	382
Problems and Questions	383
Mathematical Problems	383
Historical Questions	384
Questions for Reflection	384
34. Consolidation of the Calculus	386
34.1. Ordinary Differential Equations	387
34.1.1. A Digression on Time	389
34.2. Partial Differential Equations	390
34.3. Calculus of Variations	391
34.3.1. Euler	393
34.3.2. Lagrange	394
34.3.3. Second-Variation Tests for Maxima and Minima	394
34.3.4. Jacobi: Sufficiency Criteria	395
34.3.5. Weierstrass and his School	395
34.4. Foundations of the Calculus	397
34.4.1. Lagrange's Algebraic Analysis	398
34.4.2. Cauchy's Calculus	398
Problems and Questions	399
Mathematical Problems	399
Historical Questions	400
Questions for Reflection	400
PART VII. SPECIAL TOPICS	
Contents of Part VII	404
35. Women Mathematicians	405
35.1. Sof'ya Kovalevskaya	406
35.1.1. Resistance from Conservatives	408

35.2. Grace Chisholm Young	408
35.3. Emmy Noether	411
Questions	414
Historical Questions	414
Questions for Reflection	415
36. Probability	417
36.1. Cardano	418
36.2. Fermat and Pascal	419
36.3. Huygens	420
36.4. Leibniz	420
36.5. The <i>Ars Conjectandi</i> of James Bernoulli	421
36.5.1. The Law of Large Numbers	422
36.6. De Moivre	423
36.7. The Petersburg Paradox	424
36.8. Laplace	425
36.9. Legendre	426
36.10. Gauss	426
36.11. Philosophical Issues	427
36.12. Large Numbers and Limit Theorems	428
Problems and Questions	429
Mathematical Problems	429
Historical Questions	430
Questions for Reflection	431
37. Algebra from 1600 to 1850	433
37.1. Theory of Equations	433
37.1.1. Albert Girard	434
37.1.2. Tschirnhaus Transformations	434
37.1.3. Newton, Leibniz, and the Bernoullis	436
37.2. Euler, D'Alembert, and Lagrange	437
37.2.1. Euler	437
37.2.2. D'Alembert	438
37.2.3. Lagrange	438
37.3. The Fundamental Theorem of Algebra and Solution by Radicals	439
37.3.1. Ruffini	440
37.3.2. Cauchy	441
37.3.3. Abel	442
37.3.4. Galois	443
Problems and Questions	445
Mathematical Problems	445
Historical Questions	446
Questions for Reflection	446

38. Projective and Algebraic Geometry and Topology	448
38.1. Projective Geometry	448
38.1.1. Newton's Degree-Preserving Mapping	448
38.1.2. Brianchon	449
38.1.3. Monge and his School	450
38.1.4. Steiner	451
38.1.5. Möbius	452
38.2. Algebraic Geometry	453
38.2.1. Plücker	454
38.2.2. Cayley	455
38.3. Topology	456
38.3.1. Combinatorial Topology	456
38.3.2. Riemann	457
38.3.3. Möbius	458
38.3.4. Poincaré's <i>Analysis situs</i>	459
38.3.5. Point-Set Topology	461
Problems and Questions	462
Mathematical Problems	462
Historical Questions	463
Questions for Reflection	463
39. Differential Geometry	464
39.1. Plane Curves	464
39.1.1. Huygens	464
39.1.2. Newton	466
39.1.3. Leibniz	467
39.2. The Eighteenth Century: Surfaces	468
39.2.1. Euler	468
39.2.2. Lagrange	469
39.3. Space Curves: The French Geometers	469
39.4. Gauss: Geodesics and Developable Surfaces	469
39.4.1. Further Work by Gauss	472
39.5. The French and British Geometers	473
39.6. Grassmann and Riemann: Manifolds	473
39.6.1. Grassmann	474
39.6.2. Riemann	474
39.7. Differential Geometry and Physics	476
39.8. The Italian Geometers	477
39.8.1. Ricci's Absolute Differential Calculus	478
Problems and Questions	479
Mathematical Problems	479
Historical Questions	479
Questions for Reflection	479

40. Non-Euclidean Geometry	481
40.1. Saccheri	482
40.2. Lambert and Legendre	484
40.3. Gauss	485
40.4. The First Treatises	486
40.5. Lobachevskii's Geometry	487
40.6. János Bolyai	489
40.7. The Reception of Non-Euclidean Geometry	489
40.8. Foundations of Geometry	491
Problems and Questions	492
Mathematical Problems	492
Historical Questions	493
Questions for Reflection	493
41. Complex Analysis	495
41.1. Imaginary and Complex Numbers	495
41.1.1. Wallis	497
41.1.2. Wessel	498
41.1.3. Argand	499
41.2. Analytic Function Theory	500
41.2.1. Algebraic Integrals	500
41.2.2. Legendre, Jacobi, and Abel	502
41.2.3. Theta Functions	504
41.2.4. Cauchy	504
41.2.5. Riemann	506
41.2.6. Weierstrass	507
41.3. Comparison of the Three Approaches	508
Problems and Questions	508
Mathematical Problems	508
Historical Questions	509
Questions for Reflection	509
42. Real Numbers, Series, and Integrals	511
42.1. Fourier Series, Functions, and Integrals	512
42.1.1. The Definition of a Function	513
42.2. Fourier Series	514
42.2.1. Sturm–Liouville Problems	515
42.3. Fourier Integrals	516
42.4. General Trigonometric Series	518
Problems and Questions	519
Mathematical Problems	519
Historical Questions	519
Questions for Reflection	519

43. Foundations of Real Analysis	521
43.1. What is a Real Number?	521
43.1.1. The Arithmetization of the Real Numbers	523
43.2. Completeness of the Real Numbers	525
43.3. Uniform Convergence and Continuity	525
43.4. General Integrals and Discontinuous Functions	526
43.5. The Abstract and the Concrete	527
43.5.1. Absolute Continuity	528
43.5.2. Taming the Abstract	528
43.6. Discontinuity as a Positive Property	529
Problems and Questions	530
Mathematical Problems	530
Historical Questions	531
Questions for Reflection	531
44. Set Theory	532
44.1. Technical Background	532
44.2. Cantor's Work on Trigonometric Series	533
44.2.1. Ordinal Numbers	533
44.2.2. Cardinal Numbers	534
44.3. The Reception of Set Theory	536
44.3.1. Cantor and Kronecker	537
44.4. Existence and the Axiom of Choice	537
Problems and Questions	540
Mathematical Problems	540
Historical Questions	541
Questions for Reflection	541
45. Logic	542
45.1. From Algebra to Logic	542
45.2. Symbolic Calculus	545
45.3. Boole's <i>Mathematical Analysis of Logic</i>	546
45.3.1. Logic and Classes	546
45.4. Boole's <i>Laws of Thought</i>	547
45.5. Jevons	548
45.6. Philosophies of Mathematics	548
45.6.1. Paradoxes	549
45.6.2. Formalism	550
45.6.3. Intuitionism	551
45.6.4. Mathematical Practice	553

45.7. Doubts About Formalized Mathematics: Gödel's Theorems	554
Problems and Questions	555
Mathematical Problems	555
Historical Questions	555
Questions for Reflection	556
Literature	559
Name Index	575
Subject Index	585

PREFACE

Like its immediate predecessor, this third edition of *The History of Mathematics: A Brief Course* must begin with a few words of explanation to users of the earlier editions. The present volume, although it retains most of the material from the second edition, has been reorganized once again. In the first edition each chapter was devoted to a single culture or period within a single culture and subdivided by mathematical topics. In the second edition, after a general survey of mathematics and mathematical practice in Part I, the primary division was by subject matter: numbers, geometry, algebra, analysis, mathematical inference. After long consideration, I found this organization less desirable than a chronological ordering. As I said in the preface to the second edition,

For reasons that mathematics can illustrate very well, writing the history of mathematics is a nearly impossible task. To get a proper orientation for any particular event in mathematical history, it is necessary to take account of three independent “coordinates”: the time, the mathematical subject, and the culture. To thread a narrative that is to be read linearly through this three-dimensional array of events is like drawing one of Peano’s space-filling curves. Some points on the curve are infinitely distant from one another, and the curve must pass through some points many times. From the point of view of a reader whose time is valuable, these features constitute a glaring defect. The problem is an old one, well expressed eighty years ago by Felix Klein, in Chapter 6 of his *Lectures on the Development of Mathematics in the Nineteenth Century*:

I have now mentioned a large number of more or less famous names, all closely connected with Riemann. They can become more than a mere list only if we look into the literature associated with the names, or rather, with those who bear the names. One must learn how to grasp the main lines of the many connections in our science out of the enormous available mass of printed matter without getting lost in the time-consuming discussion of every detail, but also without falling into superficiality and dilettantism.

I have decided that in the lexicographic ordering of the three-dimensional coordinate system mentioned above, culture is the first coordinate, chronology the second, and mathematical content the third. That is the principle on which the first six parts of the present edition are organized. In the seventh and final part, which covers the period from 1800 on, the first coordinate becomes irrelevant, as mathematics acquires a worldwide scope. Because so much new mathematics was being invented, it also becomes impossible to give any coherent description of its whole over even a single decade, and so the chronological ordering has to become the second coordinate, as mathematical content becomes the first.

Changes from the Second Edition

Besides the general reorganization of material mentioned above, I have also had a feeling that in the previous edition I succumbed in too many places to the mathematician’s impulse

to go into mathematical detail at the expense of the history of the subject and to discuss some questions of historical minutiae that are best omitted in a first course. I have therefore condensed the book somewhat. The main difference with earlier editions is that I have tried to adapt the text better to the needs of instructors. To that end, I have made the chapters more nearly uniform in length, usually ten to twelve pages each, putting into each chapter an amount of material that I consider reasonable for a typical 50-minute class. In addition, I have scrutinized the problems to be sure that they are reasonable as homework problems. They are of three types: (1) those that develop a mathematical technique, such as the Chinese method of solving polynomial equations numerically, the *kuttaka*, computation by the Egyptian method, *prosthaphæresis*, and the like; (2) those that ask the student to recall a specific set of historic facts (these generally have brief answers of a sentence or two and should be answerable directly from the narrative); and (3) those that ask the student to speculate and synthesize the history into a plausible narrative, including possible motives for certain investigations undertaken by mathematicians. In survey chapters at the beginning of some parts, only the last two types occur.

The book is divided into seven parts. The first six, comprising the first 34 chapters, contain as systematic a discussion as I can manage of the general history of mathematics up to the nineteenth century. Because it is aimed at a general audience, I have given extra attention to topics that continue to be in the school curriculum, while at the same time trying to discuss each topic within the context of its own time. At the end of each chapter are a few questions to provide a basis for classroom discussions. More such questions can be found in the accompanying teacher's manual. I believe that these 34 chapters, totaling about 400 pages, constitute a one-semester course and that any extra class meetings (I assume 42 such meetings) will be devoted to quizzes, midterms, and perhaps one or two of the specialized chapters in Part VII.

The seventh and last part of the book consists of more narrowly focused discussions. Except for Chapter 35, which discusses a small portion of the history of women in mathematics, these are updates, arranged by subject matter and carrying the history of the topics they treat into the twentieth century. Since this material involves modern mathematics, it is technically much more difficult than the first six parts of the book, and the mathematical homework problems reflect this greater difficulty, making much higher demands on the reader's mathematical preparation. Instructors will of course use their own judgment as to the mathematical level of their students. In some of these chapters, I have exceeded the self-imposed limit of 12 pages that I tried to adhere to in the first six parts of the book, assuming that instructors who wish to discuss one of these chapters will be willing to devote more than one class meeting to it.

Elementary Texts on the History of Mathematics

A textbook on the history of mathematics aimed at a first course in the subject, whose audience consists of teachers, mathematicians, and interested students from other specialties, cannot be as complete or as focused as an encyclopedia of the subject. Connections with other areas of science deserve attention quite as much as historical issues of transmission and innovation. In addition, there are many mathematical skills that the reader cannot be presumed to have, and these need to be explained as simply as possible, even when the explanation does not faithfully reproduce the historical text in which the subject arose. Thus, I have hybridized and simplified certain mathematical techniques in order to provide a usable model of what was actually done while stripping away complications that make

the original texts obscure. This much sacrifice of historical accuracy is necessary, I believe, in order to get to the point within the confines of a single semester. At the same time, I think the exposition of these and other topics gives a reasonable approximation to the essence of the original texts.

This concept of a reasonable approximation to the original presents a problem that requires some judgment to solve: How “authentic” should we be when discussing works written long ago and far away, using concepts that have either disappeared or evolved into something very different? Historians have worked out ways of giving some idea of what original documents looked like. We can simply write numbers, for example, in our own notation. But when those numbers are part of a system with operational connections, it is necessary to invent something that is isomorphic to the original system, so that, for example, numbers written in sexagesimal notation still have a sexagesimal appearance, and computations done in the Egyptian manner are not simply run through a calculator and the output used. This problem is particularly acute in Euclidean geometry, which makes no reference to any units of length, area, or volume. The “Euclidean” geometry that is taught to students in high school nowadays freely introduces such units and makes use of algebraic notation to give formulas for the areas and volumes of circles, spheres, cones, and the like. This modernization conceals the essence of Euclid’s method, especially his theory of proportion. He did not speak of the area of a circle, for example, only of the ratio of one circle to another, proving that it was the same as the ratio of the squares having their diameters as sides (Book 12, Proposition 2). How much of that authentic Euclidean geometry, which I call *metric-free*, should the student be subjected to? Without it, many of the most important theorems proved by Euclid, Archimedes, and Apollonius look very different from their original forms. On the other hand, it *is* cumbersome to expound, and one is constantly tempted to capitulate and “modernize” the discussion. I have made the decision in this book to draw the line at conic sections, using symbolic notation to describe them, though I do so with a very bad conscience. But I would never dream of presenting, in an introductory text, the actual definition of the *latus rectum* given by Apollonius. I try to hold the use of symbolic algebra to a minimum, but compromises are necessary in the real world.

When it comes to algebra, symbolic notation is a very late arrival. Algorithms for solving cubic and quartic equations preceded it, and those algorithms are very cumbersome to explain without symbols. Once again, I surrender to necessity and try to present the essence of the method without getting bogged down in the technical details of the original works. There is a further difficulty that most students have learned algebra by rote and can carry out certain operations, but have no insight into the essence of the problems they have been taught to solve. They may know what American students call the FOIL method of solving quadratic equations with integer coefficients, and some of them may even remember the quadratic formula, but I have yet to encounter a student who has grasped the simple fact that solving a quadratic equation is a way of finding two numbers if one knows their sum and product. Nor have I found a student who has the more general insight that classical algebra is the search for ways of rendering explicit numbers that are determined only implicitly, even though this insight is crucial for recognizing algebra when it occurs in early treatises, where there is no symbolic notation.

Besides the enormous amount of mathematics that the human race has created, so enormous that no one can be really expert except over a tiny region of it, the historian has the additional handicap of trying to fit that mathematics into the context of a wide range of cultures, most of which will not be his or her area of expertise. I feel these limitations with

particular keenness when it comes to languages. Despite a lifetime spent trying to acquire new languages in what spare time I have had, I really feel comfortable (outside of English, of course) only when working in Russian, French, German, Latin, and ancient Greek. (I have acquired only a modest ability to read a bit of Japanese, which I constantly seek to expand.) Of course, having a language from each of the Romance, Germanic, and Slavic groups makes it feasible to attempt reading texts in perhaps two dozen languages, but one needs to be on guard and never rely on one's own translations in such cases. I am most sharply aware of my total dependence on translations of works written in Chinese, Sanskrit, and Arabic. Even though I report what others have said about certain features of these languages for the reader's information, let it be noted here and now that anything I say about any of these languages is pure hearsay.

Just to reiterate: One can really glimpse only a small portion of the history of mathematics in an introductory course. Some idea of how much is being omitted can be seen by a glance at the website at the University of St Andrews.

<http://www-history.mcs.st-and.ac.uk/>

That site provides biographies of thousands of mathematicians. Under the letter G alone there are 125 names, fewer than 40 of which appear in this book. While many of the "small fry" have made important contributions to mathematics, they do not loom large enough to appear on a map the size and scale of the present work. Thus, it needs to be kept in mind that the picture is being painted in very broad brush strokes, and many important details are simply not being shown. Every omission is regrettable, but omissions are necessary if the book is to be kept within 600 pages.

And, finally, a word about the cover. When I was asked what kind of design I wished, I thought of a collage of images encompassing the whole history of the subject: formulas and figures. In the end, I decided to keep it simple and let one part stand for the whole. The part I chose was the conic sections, because of the length and breadth of their influence on the history of the subject. Arising originally as tools to solve the problems of trisecting the angle and doubling the cube, they were the subject of one of the profoundest treatises of ancient times, that of Apollonius. Later, they turned out to be the key to solving cubic and quartic equations in the work of Omar Khayyam, and they became a laboratory for the pioneers of analytic geometry and calculus to use in illustrating their theories. Still later, they were a central topic in the study of projective geometry, and remained so in algebraic geometry far into the nineteenth century. It is no accident that non-Euclidean geometries are classified as elliptic and hyperbolic, or that linear partial differential equations are classified as elliptic, parabolic, and hyperbolic. The structure revealed by this trichotomy of cases for the intersection of a plane with a cone has been enormous. If any one part deserves to stand for the whole, it is the conic sections.

WHAT IS MATHEMATICS?

This first part of our history is concerned with the “front end” of mathematics (to use an image from computer algebra)—its relation to the physical world and human society. It contains some general considerations about mathematics, what it consists of, and how it may have arisen. This material is intended as an orientation for the main part of the book, where we discuss how mathematics has developed in various cultures around the world. Because of the large number of cultures that exist, a considerable paring down of the available material is necessary. We are forced to choose a few sample cultures to represent the whole, and we choose those that have the best-recorded mathematical history. The general topics studied in this part involve philosophical and social questions, which are themselves specialized subjects of study, to which a large amount of scholarly literature has been devoted. Our approach here is the naive commonsense approach of an author who is not a specialist in either philosophy or sociology. Since present-day governments have to formulate *policies* relating to mathematics and science, it is important that such questions not be left to specialists. The rest of us, as citizens of a republic, should read as much as time permits of what the specialists have to say and make up our own minds when it comes time to judge the effects of a policy.

Contents of Part I

1. Chapter 1 (Mathematics and Its History) considers the general nature of mathematics and gives an example of the way it can help to understand the physical world. We also outline a series of questions to be kept in mind as the rest of the book is studied, questions to help the reader flesh out the bare bones in the historical documents.
2. Chapter 2 (Proto-mathematics) studies the mathematical reasoning invented by people in the course of solving the immediate and relatively simple practical problems of administering a government or managing a construction site. In this area we are dependent on archaeologists and anthropologists for the historical information available.

Mathematics and its History

...all histories, to the extent that they contain a system, a drama, or a moral, are so much literary fiction.

Those who cannot remember the past are condemned to repeat it. (Often misquoted as “Those who do not learn from history are doomed to repeat it.”)

George Santayana (1863–1952), Spanish–American philosopher
(born Jorge Agustín Nicolás Ruiz de Santayana y Borrás)

The history of mathematics is a hybrid subject, taking its material from mathematics and history, sometimes invoking other areas such as psychology, political history, sociology, and philosophy to give a detailed picture of the development of mathematics. Obviously, no one can be an expert in all of these areas, and some compromises have to be accepted. Especially in an introductory course, it is often necessary to oversimplify both the mathematics itself and the social and historical context in which it arose so that the most significant portions can be included. No history of the subject that covers more than a narrow band of time can aim for anything like completeness.

1.1. TWO WAYS TO LOOK AT THE HISTORY OF MATHEMATICS

One of the most distinguished historians of mathematics, Ivor Grattan-Guinness (b. 1941), has made a distinction between *history* and *heritage*. History asks the question “What happened in the past?” Heritage asks “How did things come to be the way they are?” Obviously, the first of these two questions is more general than the second. Many things happened in the past that had no influence on the current shape of things, not only in mathematics but in all areas of human endeavor, including art, music, and politics. Such events are history, but not heritage. The study of history in this sense is a purely intellectual exercise, not aimed at any applications, nor to teach a moral, nor to make people better citizens. What it does aim at is getting an accurate picture of the past for the edification of those who have a taste for such knowledge. It is difficult to write such a history, as the first epigram from George Santayana given above shows.

Even on the most impersonal, objective level, we don’t want the raw, unedited past, which is a raging tsunami of sneezes and hiccups; some judgment is needed to select the events in

the past that are of interest. To that extent, Santayana's implication is correct: All history is literary fiction. The danger for the historian lies in trying to frame a particular picture of the past in order to make it tell the story that one personally would like to hear. In the history of mathematics, there is a special danger because the mathematics itself fits together in a very logical way, while the routes by which it has been discovered and developed have all the illogical disorder that is inherent in any process involving human thinking. For example, it is known that there is no finite algebraic formula involving only arithmetic operations and root extractions that will yield a root of every quintic equation when the coefficients of that equation are substituted for its variables. This result follows very neatly from what we now call Galois theory, after Evariste Galois (1811–1832), who first introduced its basic ideas. It is nowadays always proved using this technique. But the theorem was first stated and given a semblance of a proof by Paolo Ruffini (1765–1822) and Niels Henrik Abel (1802–1829), neither of whom knew Galois theory. They both proceeded by counting the number of different values that such a hypothetical formula would generate if all possible values were substituted in the formula for each n th root it contains. This example is typical of many cases in the history of mathematics, where the proof of a proposition resulted not from rigorously arranged steps following in logical order from one another, but from a number of independent ideas gradually coming into focus.

1.1.1. History, but not Heritage

During the fifteenth and sixteenth centuries, tables of sines were used to simplify multiplication and reduce it to addition and subtraction. This procedure was called *prosthaphæresis*, from the Greek words *prosthairesis* (προσθαίρεσις), meaning *taking toward*, and *aphairesis* (ἀφαίρεσις), meaning *taking away*. This technique disappeared almost without a trace after the discovery of logarithms in the early seventeenth century, and it is nowadays unknown even to most professional mathematicians. Nevertheless, it was an important idea in its time and deserves to be remembered. We shall take the time to discuss it and practice it a bit. As we shall see, it is actually more efficient than logarithms for computing the formulas of spherical trigonometry.

1.1.2. Our Mathematical Heritage

The appeal of history is to a person of a particular “antiquarian” bent of mind. Heritage, which is parasitic upon history, has a somewhat more practical aim: to help us understand the world that we ourselves live in. This is the “useful” part of history that historians advertise to the public to gain support, and it is the point of view expressed in the second of Santayana's epigrams at the beginning of this lecture. (Notice that the two epigrams taken together imply that the human race needs a variety of history that is actually literary fiction.)

If you have taken a course called “modern algebra,” for example, you found yourself confronted with a collection of abstract objects—groups, rings, fields, vector spaces—that seemed to have nothing in common with high-school algebra except that they required the use of letters. How did these abstract subjects come to be referred to as algebra? By tracing the story of the unsolvability of the general equation of degree five, we can answer this question.

After algebraic formulas were found for solving equations of degree 3 and 4 in the sixteenth century, two centuries were spent in the quest for a mathematical “Holy Grail,” an

algebraic formula to solve the general equation of degree 5. Some people thought they had succeeded; but in the late eighteenth and early nineteenth centuries, Ruffini, Abel (one of those who for a time thought he had succeeded in finding the formula), and William Rowan Hamilton (1805–1865) were able to show that no such formula could exist. The question then arose of determining which equations *could* be solved by algebraic operations (the operations of arithmetic, together with the extraction of roots) and which could not. The answer to this question, as shown by Galois, depends on the abstract nature of a certain set of permutations of the roots. This was the beginning of the study of groups, a word first used by Galois. The concept of an abstract group arose some decades later, along with the rest of these abstract creations, all of which found numerous applications in other areas of mathematics. The original problem that gave rise to much of this modern algebra was, in the end, only one part of the vast edifice of modern algebra.

1.2. THE ORIGIN OF MATHEMATICS

The farther we delve into the past, the more we find mathematics entangled with accounting, surveying, astronomy, and the general administration of empires. Mathematics arises wherever people think about the physical world or about the world of ideas embodied in laws and even theology. It grows like a plant, from a seed that germinates and later ramifies to produce roots, branches, leaves, flowers, and fruit. It is constantly growing.

1.2.1. Number

It seems nearly certain that the small positive integers, the kinds of numbers that are intuitively known to everyone, are the “seed” of mathematics. Essentially all mathematical concepts can be traced ultimately to the use of numbers to explain the world. Numbers seem to be a universal mode of human thought. They were probably used originally in a kind of informal accounting, when it was necessary to keep track of objects that could be regarded as interchangeable, such as the cattle in a herd. Through anthropology, archaeology, and written texts, we can trace a general picture of arithmetical progress in handling such discrete collections, from counting, through computation, and finally to abstract number theory. Many different cultures have shown a convergent development in this area, although in the final stage there is considerable variety in the choice of topics developed. Through this history, we shall gradually introduce the properties of numbers in the chapters that follow. At the moment, we take note of just one important property that they have, namely that discrete collections can be *exactly* equal: If I have \$9845.63 in my checking account, and you have \$9845.63 in your checking account, then we have *exactly* the same amount of money, for all financial purposes whatsoever. When you count—votes, pennies, or attendance at a football game—it is at least theoretically possible to get the outcome exactly right, with no error at all.

1.2.2. Space

While discrete collections are naturally handled through counting, nature presents us with the need to measure quantities that are continuous rather than discrete, quantities such as

length, area, volume, weight, time, and speed. Number is invoked to solve such problems; but in each case, it is necessary to *choose a unit* and regard the continuous quantity as if it were a discrete collection of units. Doing so adds a layer of complication, since the unit is arbitrary and culturally dependent. When people from different groups meet and talk about such quantities, they need to reconcile their units.

The essence of continuous quantities is that they can be divided into pieces of arbitrarily small size. A continuous measurement therefore always has precision limited by the size of the unit chosen. Equality of two continuous objects of the same kind is always approximate, only up to the standard unit of measurement in which their sizes are expressed.¹

This distinction between the discrete and the continuous is fundamental in mathematics, and it brings with it many metaphysical and mathematical complications. In particular, the notion of infinite precision, which is required to define what we call “real numbers,” is difficult to visualize and define. In “real-world” applications, computers finesse the problem by replacing real numbers with floating-point numbers. The result is that most engineers and scientists never really have to encounter the difference between the two, and mathematicians who attempt to talk to them tend to forget that they are not really speaking the same language. Some computer algebra programs (*Mathematica* and *Maple*, for example) will handle irrational numbers like π and $\sqrt{2}$ symbolically and will convert them to decimal approximations only when a numerical result is requested by the user.

The ideas needed to handle continuous quantities were first applied to lengths, areas, and volumes. They led to geometry, which arose in many different places as a way of comparing the sizes of objects having different shapes. Like the positive integers, these shapes seem to be a cultural universal, as all over the world we find people discussing triangles, squares, rectangles, circles, spheres, pyramids, and the like. Moreover, the shape of a standard unit area is universally square.

As happens with arithmetic, geometry passes through certain stages in a particular order in many different cultures. The first stage is simply measurement, finding physical ways of counting how many standard units of length, area, or volume there are in a piece of rope, a plot of land, or a ditch that is to be excavated. Soon, the processes of arithmetic are invoked to provide indirect ways of computing areas and volumes. At this stage, the universal shapes named above are isolated for study. Finally, relations among the parts of a geometric figure are studied, leading to abstract geometry, and some way is found to give geometric demonstrations of relations that are not obvious. Here again we find a cultural universal in the Pythagorean theorem, which was apparently discovered in several places independently. Since it invokes the notion of a square, it shows that human imagination is by nature Euclidean.² This third stage occurs in several places, among them Mesopotamia, China, India, and ancient Greece. In addition, the ancient Greeks mixed philosophy and abstract logic into their geometry and number theory, producing a number of long treatises that were unique in their time and became a model for later mathematical writing the world over.

¹We may or may not have in the back of our minds a picture of an infinitely precise number that represents the *exact* volume of water in a jar, for example, but we can meaningfully *talk about* only a *measured* volume and say with absolute assurance that the “true” volume lies between two limits. This “true” volume, given with infinite precision, is unknowable. This problem always arises in applications to the physical world. It is meaningful to ask what the 3000th decimal digit of $\frac{1}{\pi}$ is (it is 2); it does not make sense to ask what the 3000th decimal digit of Planck’s constant in MKS units is.

²Rectangles do not exist in non-Euclidean geometry. There is a Pythagorean theorem in both elliptic and hyperbolic geometry, but it involves trigonometric and hyperbolic functions and is more analytic than geometric in nature.

1.2.3. Are Mathematical Ideas Innate?

The cross-cultural constancy of the arithmetic operations and the standard shapes of geometry is a very striking fact and lends some support to the views of the eighteenth-century philosopher Immanuel Kant (1724–1804), who thought mathematical knowledge was “synthetic but *a priori*.” By that he meant that statements such as $7 + 5 = 12$ or that a triangle can be constructed having sides of three given lengths provided the sum of the smaller two exceeds the third (his examples) are not things we learn from observation, like that statement that pandas eat bamboo. Things learned from experience are *a posteriori* in Kant’s language.

Mathematical facts are, as we would now say, “hard-wired” into the human brain, or, as Kant said, *a priori* (anterior to experience). At the same time, they are “synthetic,” that is, they are not mere tautologies like the statement that two first cousins have a common grandparent. Tautological statements were called *analytic* by Kant, meaning that the very definition of first cousinhood involves a common grandparent, but (Kant said) the notions of 7, 5, and addition do not by their nature involve the number 12. We shall return to this topic when we discuss logic in Chapter 45.

In the late nineteenth and early twentieth century, a school of philosophy of mathematics arose known as logicism. Its adherents defended the proposition that mathematics could be derived from logic. If they are correct, then Kant’s belief that arithmetic and geometric propositions are synthetic must be wrong. It is true that logicists produce a formal proof of the simple fact that twice two make four. In that sense, they have *made* this proposition analytic rather than synthetic. Still, there is a more colloquial sense of the word *proof* that is violated in the process. A proof is usually thought of as deriving a proposition that is *not* obvious from others that *are* obvious. Unfortunately, the axioms of set theory are very far from being more obvious than the equality $2 + 2 = 4$.

1.2.4. Symbolic Notation

We noted above that symbolism entered mathematics via algebra, as the most elegant way of giving a description of an unknown or unspecified number. Eventually, this symbolism conquered number theory and geometry as well, and there is now no branch of mathematics, pure or applied, that is not dominated by symbolic formulas. This tool for thinking is so important that we shall consider it a third ingredient of mathematics, after number and space. Algebra itself, however, got along without symbols for centuries. If algebra is defined as a subject where symbols are used to represent unspecified numbers in equations, then we shall find no algebra at all until a few centuries ago. But the essence of algebra is not in the symbolism, or even in the equation. It is in the process of *naming* an implicitly defined number, and that will be our definition. Thus, finding a number that yields 24 when squared and added to five times itself is an algebra problem. It need not be stated as the equation $x^2 + 5x = 24$.

1.2.5. Logical Relations

The fourth and last ingredient of mathematics is the logical organization of the subject. The strict formalism that we now associate with mathematical theories was first set out in connection with geometry and number theory in ancient Greece. The earliest major work embodying it is Euclid’s *Elements*, which was the model for later work by Greek

mathematicians such as Archimedes, Apollonius, Ptolemy, Pappus, and others and became the inspiration for many of the classic treatises of modern mathematics, such as Newton's famous *Philosophiæ naturalis principia mathematica* (*Mathematical principles of natural philosophy*).³

1.2.6. The Components of Mathematics

We shall consider mathematics as being made up of the four basic components just described. The first of these can be loosely described as arithmetic,⁴ the second as geometry, the third as algebra, and the fourth as mathematical reasoning.

Out of these four elements arise calculus, probability, statistics, set theory, topology, complex analysis, mathematical logic, and a host of other areas of modern mathematics that make it the magnificent monument to the human intellect that it is.

1.3. THE PHILOSOPHY OF MATHEMATICS

If we were to study merely what happened in the past, even if we did it with an eye toward the present, the development of mathematics would seem very much like one wave after another breaking on the shore. Without some interesting conjectures as to what the creators were trying to do, it would be difficult to make any sense of this history. This problem is particularly acute in algebra, as you may recall from the “story problems” you were asked to solve in high school, which are without exception, colossally useless. Surely the complicated mathematical reasoning in this subject was not invented in order to find out when two trains will meet if they set out from different stations at different times. In order to flesh out the subject and paint it in brighter and more realistic colors, we need to ask ourselves broad philosophical questions while we are studying the past. Here is a short list of questions of interest.

Epistemological Questions (Theory of Knowledge)

1. What is the nature of mathematical objects such as numbers, triangles, probabilities, and functions? In what sense do they “exist”?
2. What can we know about infinite collections of things? Is a finite human mind capable of knowing infinitely many different things?
3. What is meant by continuity? Is it possible to formulate continuity in the discrete symbols of ordinary language?

Metaphysical Questions (Nature of Reality)

1. What is the relation of the objects of pure mathematics to those of applied mathematics? Many logical relations exist in pure mathematics; and when they are applied in

³Natural philosophy was the name once given to what we now call the natural sciences.

⁴Throughout most of history, however, *arithmetic* meant what we now call *number theory*. The modern use of this word to denote the four basic operations on numbers is largely an American innovation, and not a desirable one, in my opinion. Since the word comes from Greek and uses the root *arithmos*, meaning *number*, and the suffix *-ikos*, meaning *skilled in*, I prefer to translate it as *the numerical art*.

real-world situations, they very often make predictions that can be verified by observation. Does that mean that pure-mathematics relations correspond to relations in the physical universe? If so, what is the nature of these physical relations? To paraphrase the Nobel Prize-winning physicist Eugene Wigner (1902–1995), what is the reason for “the unreasonable effectiveness of mathematics” in explaining the world? For a republication of Wigner’s 1960 paper on this subject, see the following website:

<http://www.dartmouth.edu/~matc/MathDrama/reading/Wigner.html>

2. Why do probability and statistics work so well in practice that insurance companies and gambling casinos can rely on the seeming chaos of random events to stay predictable “in the large”? For that matter, why do we make the assumption that the future will resemble the past, so that we can make mathematical predictions about the future state of the physical universe?

Metamathematical Questions

1. The business of the pure mathematician is to prove theorems, that is, to make valid inferences from premises. What premises should be allowed, and what rules of inference can be trusted? There is a school of mathematicians—the intuitionists—that refuses to use certain basic logical and mathematical assumptions, chiefly the law of excluded middle (if not- A is false, then A is true) or the axiom of choice, which says intuitively that if you have a collection of containers and each one has something inside it, you can reach in and take one object out of each. The collections obtained in this way are the elements of a new set called the *Cartesian product* of the containers in the original collection.) Most mathematicians use these two principles freely and have no qualms about doing so.
2. How important is a formal, deductive presentation of a mathematical subject? Can a mathematical paper that appeals to intuition rather than formal proof be accepted as valid?

Sociological Questions

1. How important is mathematics to society? What genuine material or moral progress in the world can be traced to the activity of mathematicians?
2. What mathematics, if any, should be taught to every citizen of a modern democracy?

1.3.1. Mathematical Analysis of a Real-World Problem

We shall illustrate just one of the ways in which this course is intended to make you think about the mathematical link between the physical world around us and our thinking processes. We choose music as an example. On April 17, 1712, the philosopher–mathematician Leibniz (1646–1716) wrote to Christian Goldbach (1690–1764)

Musica est exercitium arithmeticae occultum nescientis se numerare animi. (Music is a mysterious practicing of the numerical art by a mind that does not realize it is counting.)

(See *Epistolae ad diversos*, edited by Christian Kortholt, Vol. 1, Leipzig, 1734, p. 241, Letter CLIV.)



Figure 1.1. A rhythm pattern compounded of two simple periodic beats.

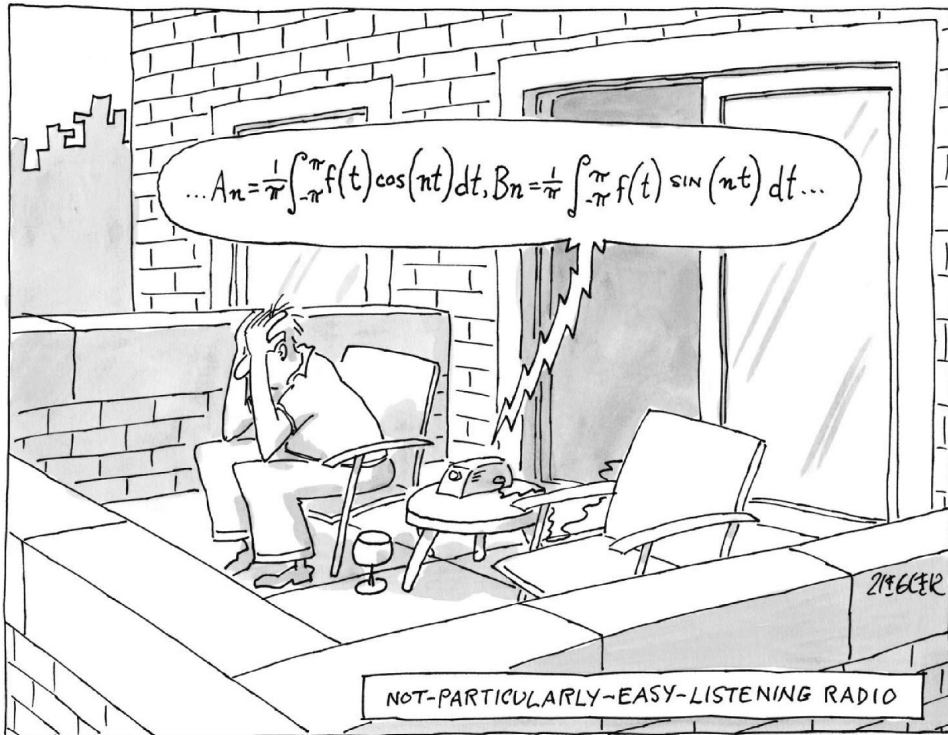
What Leibniz probably meant in this aphorism is that the rhythm patterns that are part of all music can be analyzed and found to be regular repetitions of simple periodic patterns, superimposed in some very complex ways. He may also have suspected that the pitch and quality of the notes coming from string and wind instruments can be analyzed in the same way. The rhythm is a case of a discrete phenomenon, while the pitch involves continuous periodic waves. It is possible to represent the pitch and the overtones of musical instruments as superimposed simple sine waves of various frequencies, amplitudes, and phases. From those sine waves, it is not only theoretically but also practically possible (as we now know) to synthesize music. Theoretically, one can play an entire orchestral symphony with nothing but tuning forks struck with the proper strength and at the proper times. We shall illustrate the underlying principle with a simple discrete example, leaving the more complicated continuous case for later description.

If you can read music, try tapping out the rhythm pattern depicted in Fig. 1.1. If you cannot, ask someone who can read music to do this for you.

You will find that you can duplicate this rhythm pattern if you count by twelves, bringing both hands down on the count of 1, then alternating right and left hands on 4, 5, 7, 9, and 10. In this way, the left hand can be tapping out ONE-two-three-FOUR-five-six-SEVEN-eight-nine-TEN-eleven-twelve, while the right hand is tapping ONE-two-three-four-FIVE-six-seven-eight-NINE-ten-eleven-twelve. In other words, the left hand is tapping four beats to the bar while the right hand taps three beats to the bar. Each hand is tapping a simple periodic pulse every three beats or every four beats. The combined effect is the pattern ONE-two-three-FOUR-FIVE-six-SEVEN-eight-NINE-TEN-eleven-twelve, which sounds very syncopated. Imagine this example elaborated to describe a whole orchestra, and extended to the pitch of the tones each instrument is producing, and you get an idea of the complexity of music when it is analyzed mathematically. Musical patterns, however, are *felt* by musicians; they are not produced mechanically, at least not by good musicians. The quickest way to master this rhythm pattern is simply to hear it. Nearly everyone can reproduce it, much more rapidly than anyone can count aloud, after hearing it for a few seconds. That is the point of Leibniz' comment that the mind *does not realize it is counting*. Mathematical analysis of tones and rhythms has the same relation to the pleasure of hearing music that chemical analysis of a cup of coffee has to the pleasure of drinking it.

It is not physically possible to play a symphony with tuning forks, and we don't actually do this. But we do an equivalent thing in our digital music. That is the point of the cartoon shown here, which appeared in *The New Yorker* on October 4, 2010 (p. 71).

From the mathematician's point of view, digital radio amounts to breaking the sound into a finite set of simple frequencies, each having a particular amplitude and phase. (The amplitude associated with frequency n in the cartoon is $\sqrt{A_n^2 + B_n^2}$.) When a digital radio receives that set of amplitudes and frequencies, it sends them in the form of electrical signals to the speakers, which then reproduce them as an audible signal. Since the human ear "truncates" the signal by being unable to pick up frequencies below 20 cycles per second or higher than 20,000 cycles per second, the result is what mathematicians call a band-limited signal. It is an important mathematical result—the Whittaker–Shannon interpolation



The Fourier series of a symphony. Copyright © Jack Ziegler/The New Yorker Collection.

theorem, named after Edmund Taylor Whittaker (1873–1956) and Claude Elwood Shannon (1916–2001)—that such a signal can be reproduced perfectly from a finite number of sample points.

The humor in this cartoon—imagine being given the Fourier series of a symphony and having to do the Fourier inversion in your head in order to interpret the symphony!—is rather esoteric and will be appreciated only by the tiny segment of the population that knows Fourier analysis. For the rest of the public, this cartoon was probably just one more way of saying, “Math is hard.”

1.4. OUR APPROACH TO THE HISTORY OF MATHEMATICS

We are going to study the history of mathematics partly for its intrinsic interest. That will lead us to develop a few mathematical skills that will not be of much use outside this course. We do this partly for an ethical reason: to preserve the memory of brilliant people whose contributions to human history should not be forgotten. Except for these excursions into true history, our focus is on the “heritage” aspect of mathematical history. The main aim of this course is to give insight into the way that today’s mathematics developed, the motives of its creators, and the social and intellectual context in which they worked. As Santayana said, in so doing, we are to some extent creating literary fiction. But it is useful fiction.

Questions for Reflection

At the end of most of the chapters in this book, there will be a set of mathematical problems testing for understanding of the mathematics discussed in that chapter, followed by a set of questions of historical fact to reinforce the historical narrative of the chapter, in turn followed by a set of questions calling for reflection on the historical and mathematical issues that arise in the chapter. In this introductory chapter, where we have not introduced any mathematics or discussed any systematic development of it, only the third category seems appropriate. The following questions are therefore intended to make you think about general issues such as those raised in the questions listed above.

- 1.1. In what practical contexts of everyday life are the fundamental operations of arithmetic—addition, subtraction, multiplication, and division—needed? Give at least two examples of the use of each. How do these operations apply to the problems for which the theory of proportion was invented?
- 1.2. Measuring a continuous object involves finding its ratio to some standard unit. For example, when you measure out one-third of a cup of flour in a recipe, you are choosing a quantity of flour whose ratio to the standard cup is 1 : 3. Suppose that you have a standard cup without calibrations, a second cup of unknown size, and a large bowl. How could you determine the volume of the second cup?
- 1.3. Units of time, such as a day, a month, and a year, have ratios. In fact you probably know that a year is about $365\frac{1}{4}$ days long. Imagine that you had never been taught that fact. How would you—how did people originally—determine how many days there are in a year?
- 1.4. Why is a calendar needed by an organized society? Would a very small society (consisting of, say, a few dozen families) require a calendar if it engaged mostly in hunting, fishing, and gathering vegetable food? What if the principal economic activity involved following a reindeer herd? What if it involved tending a herd of domestic animals? Finally, what if it involved planting and tending crops?
- 1.5. Describe three different ways of measuring time, based on different physical principles. Are all three ways equally applicable to all lengths of time?
- 1.6. In what sense is it possible to know the *exact* value of a number such as $\sqrt{2}$? Obviously, if a number is to be known only by its whole infinite decimal expansion, nobody does know and nobody ever will know the exact value of this number. What immediate practical consequences, if any, does this fact have? Is there any other sense in which one could be said to know this number *exactly*? If there are no direct consequences of being ignorant of its exact value, is there any practical value in having the *concept* of an exact square root of 2? Why not simply replace it by a suitable approximation such as 1.41421? Consider also other “irrational” numbers, such as π , e , and $\Phi = (1 + \sqrt{5})/2$. What is the value of having the *concept* of such numbers as opposed to approximate rational replacements for them?
- 1.7. Does the development of personal knowledge of mathematics mirror the historical development of the subject? That is, do we learn mathematical concepts as individuals in the same order in which these concepts appeared historically?

- 1.8.** Topology, which may be unfamiliar to you, studies (among other things) the mathematical properties of knots, which have been familiar to the human race at least as long as most of the subject matter of geometry. Why was such a familiar object not studied mathematically until the twentieth century?
- 1.9.** What function does logic fulfill in mathematics? Is it needed to provide a psychological feeling of confidence in a mathematical rule or assertion? Consider, for example, any simple computer program that you may have written. What really gave you confidence that it worked? Was it your logical analysis of the operations involved, or was it empirical testing on an actual computer with a large variety of different input data?
- 1.10.** Logic enters the mathematics curriculum in high-school geometry. The reason for introducing it at that stage is historical: Formal treatises with axioms, theorems, and proofs were a Greek innovation, and the Greeks were primarily geometers. There is no *logical* reason why logic is any more important in geometry than in algebra or arithmetic. Yet it seems that without the explicit statement of assumptions, the parallel postulate of Euclid would never have been questioned. Suppose things had happened that way. Does it follow that non-Euclidean geometry would never have been discovered? How important is non-Euclidean geometry, anyway? What other kinds of geometry do you know about? Is it necessary to be guided by axioms and postulates in order to discover or fully understand, say, the non-Euclidean geometry of a curved surface in Euclidean space? If it is not necessary, what is the value of an axiomatic development of such a geometry?
- 1.11.** According to musical theory, the frequency of the major fifth in each scale should be $\frac{3}{2}$ of the frequency of the base tone, while the frequency of the octave should be twice the base frequency. If you start at the lowest A on the piano and ascend in steps of a major fifth, twelve steps will bring you to the highest A on the piano. If all these fifths are tuned properly, that highest A should have a frequency of $(\frac{3}{2})^{12}$ times the frequency of the lowest A. On the other hand, that highest A is seven octaves above the lowest, so that, if all the octaves are tuned properly, the frequency should be 2^7 times as high. Now obviously, $(\frac{3}{2})^{12} \approx 129.75$ is not the same thing as $2^7 = 128$, since equality of these two quantities would mean $3^{12} = 2^{19}$, that is, an odd number would equal an even number. The difference between these two frequency ratios is called the *Pythagorean comma*. (The Greek word *komma* means a break or cutoff.) What is the significance of this discrepancy for music? Could you hear the difference between a piano tuned so that all these fifths are exactly right and a piano tuned so that all the octaves are exactly right? In fact, because of the properties of metal strings and the peculiarities of human perception, piano tuning (like music itself) is very much an art or a skill, not reducible to formula.

Proto-mathematics

Most of the history of mathematics is inferred from documents written down by scholars, starting about 4000 years ago. Before that time, and in more recent times among certain groups, mathematical ideas were being used, but not written down. In the present chapter, we shall explore this *proto-mathematics*, the kinds of mathematical thinking that people naturally engage in while going about the practical business of daily life. This inquiry assumes that there is a mode of thought called *mathematizing* that is intrinsic to human nature and therefore common to different cultures. The simplest assumption is that counting and common shapes such as squares and circles have the same meaning to everyone. Our inquiry will make use of three sources:

1. Animal behavior, which often seems to show an ability to judge numbers, shapes, and causes (if event B always follows event A, animals will come to *infer* that B is about to happen whenever A happens).
2. Developmental psychology, which reveals the order in which children develop mathematical abilities.
3. Archaeology, which sometimes turns up artifacts that seem to imply mathematical reasoning on the part of their makers.

We shall examine these sources as they relate to the four basic ingredients of mathematics listed in the preceding chapter.

2.1. NUMBER

There is ample evidence of counting from all three of the sources listed above.

2.1.1. Animals' Use of Numbers

It is not clear just how high animals and birds can count, but they certainly have the ability to distinguish not merely patterns, but actual numbers. The counting abilities of birds were studied in a series of experiments conducted in the 1930s and 1940s by O. Koehler (1889–1974) at the University of Freiburg. Koehler (1937) kept the trainer isolated from the bird. In the final tests, after the birds had been trained, the birds were filmed automatically, with

no human beings present. Koehler found that parrots and ravens could learn to compare the number of dots, up to 6, on the lid of a hopper with a “key” pattern in order to determine which hopper contained food. They could make the comparison no matter how the dots were arranged, thereby demonstrating an ability to take account of the *number* of dots rather than the *pattern*.

2.1.2. Young Children’s Use of Numbers

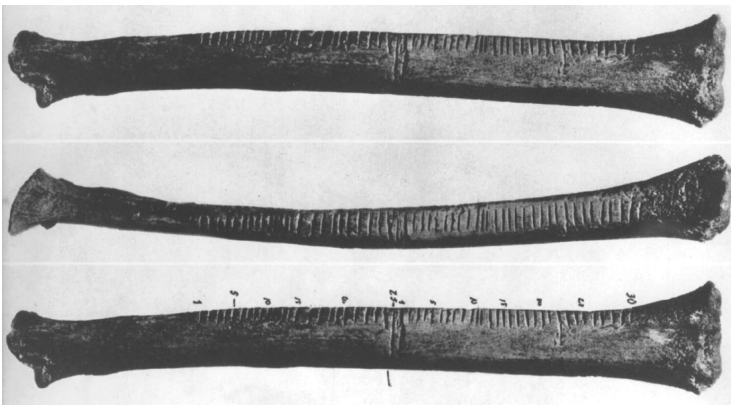
Preschool children also learn to count and use small numbers. The results of many studies have been summarized by Karen Fuson (1988). A few of the results from observation of children at play and at lessons were as follows:

1. A group of nine children aged from 21 to 45 months was found to have used the word *two* 158 times, the word *three* 47 times, the word *four* 18 times, and the word *five* 4 times.
2. The children seldom had to count “one–two” in order to use the word *two* correctly; for the word *three*, counting was necessary about half the time; for the word *four*, it was necessary most of the time; for higher numbers, it was necessary all the time.

One can thus observe in children the capacity to recognize groups of two or three without performing any conscious numerical process. This observation suggests that these numbers are primitive, while larger numbers are a conscious creation.

2.1.3. Archaeological Evidence of Counting

Animal bones containing notches have been found in Africa and Europe, suggesting that some sort of counting procedure was being carried on at a very early date, although what exactly was being counted remains unknown. One such bone, the radius bone of a wolf, was discovered at Věstonice (Moravia) in 1937. This bone is marked with two series of notches, grouped by fives, the first series containing five groups and the second six. Its discoverer,



The Věstonice wolf bone. Copyright © *Illustrated London News*, October 2, 1937. Courtesy of the Mary Evans Picture Library.

Karel Absolon (1887–1960), believed the bone to be about 30,000 years old, and modern tests on artifacts from the site appear to confirm this dating. The people who produced this bone were clearly a step above mere survival, since a human portrait carved in ivory was found in the same settlement, along with a variety of sophisticated tools. Because of the grouping by fives, it seems likely that this bone was being used to count something. Even if the groupings are meant to be purely decorative, they point to a use of numbers and counting for a practical or artistic purpose.

2.2. SHAPE

Spatial relations are also used by animals, and the gradual mastery of these phenomena by children has been charted by psychologists. In addition, many crafts, both ancient and modern show how these relations are used for both practical and decorative purposes.

2.2.1. Perception of Shape by Animals

Obviously, the ability to perceive shape is of value to an animal in determining what is or is not food, what is a predator, and so forth; and in fact the ability of animals to perceive space has been very well documented. One of the most fascinating examples is the ability of certain species of bees to communicate the direction and distance of sources of plant nectar by performing a dance inside the beehive. The pioneer in this work was Karl von Frisch (1886–1982), and his work has been continued by James L. Gould and Carol Grant Gould (1995). The experiments of von Frisch left many interpretations open and were challenged by other specialists. The Goulds performed more delicately designed experiments which confirmed the bee language by deliberately misleading the bees about the food source. The bee will traverse a circle alternately clockwise and counterclockwise if the source is nearby. If it is farther away, the alternate traversals will spread out, resulting in a figure eight, and the dance will incorporate sounds and wagging. By moving food sources, the Goulds were able to determine the precision with which this communication takes place (about 25%). Still more intriguing is the fact that the direction of the food source is indicated by the direction of the axis of the figure eight, oriented relative to the sun if there is light and relative to the vertical if there is no light.

As another example, in his famous experiments on conditioned reflexes using dogs as subjects the Russian scientist Pavlov (1849–1936) taught dogs to distinguish ellipses of very small eccentricity from circles. He began by projecting a circle of light on the wall each time he fed the dog. Eventually the dog came to expect food (as shown by salivation) every time it saw the circle. When the dog was conditioned, Pavlov began to show the dog an ellipse in which one axis was twice as long as the other. The dog soon learned not to expect food when shown the ellipse. At this point the malicious scientist began making the ellipse less eccentric and found, with fiendish precision, that when the axes were nearly equal (in a ratio of 8:9, to be exact) the poor dog had a nervous breakdown (Pavlov, 1928, p. 122).

2.2.2. Children's Concepts of Space

The most famous work on the development of mathematical concepts in children is due to Jean Piaget (1896–1980) of the University of Geneva, who wrote several books on the subject, some of which have been translated into English. Piaget divided the development

of the child's ability to perceive space into three periods: a first period (up to about 4 months of age) consisting of pure reflexes and culminating in the development of primary habits, a second period (up to about one year) beginning with the manipulation of objects and culminating in purposeful manipulation, and a third period in which the child conducts experiments and becomes able to comprehend new situations. He categorized the primitive spatial properties of objects as proximity, separation, order, enclosure, and continuity. These elements are present in greater or less degree in any spatial perception. In the baby they come together at the age of about 2 months to provide recognition of faces. The human brain seems to have some special "wiring" for recognizing faces.¹

The interesting thing about these concepts is that mathematicians recognize them as belonging to the subject of topology, an advanced branch of geometry that developed in the late nineteenth and early twentieth centuries. It is an interesting paradox that the human ability to perceive shape depends on synthesizing topological concepts; this progression reverses the pedagogical and historical ordering between geometry and topology. Piaget pointed out that children can make topological distinctions (often by running their hands over models) before they can make geometric distinctions. Discussing the perceptions of a group of 3-to-5-year-olds, Piaget and Inhelder (1967) stated that the children had no trouble distinguishing between open and closed figures, surfaces with and without holes, intertwined rings and separate rings, and so forth, whereas the seemingly simpler relationships of geometry—distinguishing a square from an ellipse, for example—were not mastered until later.

2.2.3. Geometry in Arts and Crafts

Weaving and knitting are two excellent examples of activities in which the spatial and numerical aspects of the world are combined. Even the sophisticated idea of a rectangular coordinate system is implicit in the placing of different-colored threads at intervals when weaving a carpet or blanket so that a pattern appears in the finished result.

Marcia Ascher (1991) has assembled many examples of rather sophisticated mathematics connected with arts and crafts. The Bushoong people of Zaire make part of their living by supplying embroidered cloth, articles of clothing, and works of art to others in the economy of the Kuba chiefdom. As a consequence of this work, perhaps as preparation for it, Bushoong children amuse themselves by tracing figures on the ground. The rule of the game is that a figure must be traced without repeating any strokes and without lifting the finger from the sand. In graph theory, this problem is known as the *unicursal tracing problem*. It was analyzed by the Swiss mathematician Leonhard Euler (1707–1783) in the eighteenth century in connection with the famous Königsberg bridge problem. According to Ascher, in 1905 some Bushoong children challenged the ethnologist Emil Torday (1875–1931) to trace a complicated figure without lifting his finger from the sand. Torday did not know how to do this, but he did collect several examples of such figures. The Bushoong children seem to learn intuitively what Euler proved mathematically: A unicursal tracing of a connected graph is possible if there are at most two vertices where an odd number of edges meet. The Bushoong children become very adept at finding such a tracing, even for figures as complicated as that shown in Fig. 2.1.

¹And for "seeing" faces on the moon, in clouds, and on burnt pieces of toast!

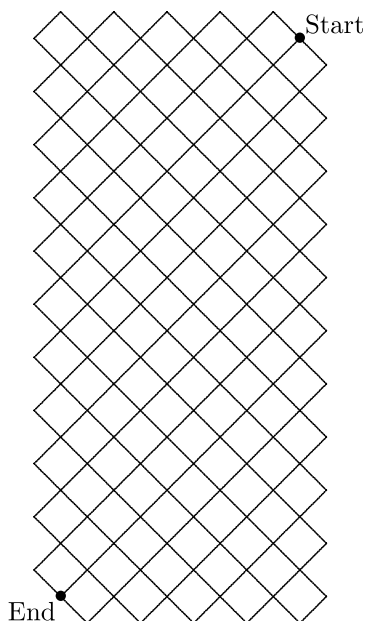


Figure 2.1. A graph for which a unicursal tracing is possible.

Examples of intricate geometric patterns susceptible to mathematical analysis are abundant throughout human history. A recent example is the book of Belcastro and Yackel (2008), which gives detailed analyses of the connections between mathematics and needlework.

2.3. SYMBOLS

Visual symbolism seems to be a peculiarly human mode of thought, not observable in animals, and one learned by children only through teaching. Artifacts from archaeological sites are sometimes interpreted as representations of divinities, but generally ancient paintings and statues tend to represent physical objects, with certain distortions in size that reflect their relative importance to the artist, or the artist's employer, in the case of Egyptian paintings that show the pharaoh much larger than anyone else.

We tend to think of symbolism as arising in algebra, since that is the subject in which we first become aware of it as a concept. The thing itself, however, is implanted in our minds much earlier, when we learn to talk. Human languages, in which sounds correspond to concepts and the temporal order or inflection of those sounds maps some relation between the concepts they signify, exemplify the process of abstraction and analogy, essential elements in mathematical reasoning.

Once numbers have been represented symbolically, the next logical step would seem to be to introduce symbols for arithmetic operations or for combining the number symbols in other ways. This step may not be necessary for rapid computation, since mechanical devices such as counting rods, pebbles, counting boards, and the like can be used as analog computers. The symbolic ability of the human mind is shown when pebbles or tally marks are used to *represent* objects in the mind of the calculator. The operations performed using these methods can rise to a high level of sophistication without the need for any written

computations. An example of the use of an automatic counting device is given by Ascher (1997) in a discussion of a system of divination used by the Malagasy of Madagascar, in which four piles of seeds are arranged in a column and the seeds removed from each pile two at a time until only one or two seeds remain. Each set of seeds in the resulting column can be interpreted as “odd” or “even.” After this procedure is performed four times, the four columns and four rows that result are combined in different pairs using the ordinary rules for adding odds and evens to generate eight more columns of four numbers. The accuracy of the generation is checked by certain mathematical consequences of the method used. If the results are satisfactory, the 16 sets of four odds and evens are used as an oracle for making decisions and ascribing causes to such events as illnesses.

Divination seems to fulfill a nearly universal human desire to feel in control of the powerful forces that threaten human happiness and prosperity. It manifests itself in a variety of ways, as just shown by the example of the Malagasy. We could also cite large parts of the Jewish *Kabbalah*, the mysticism of the Pythagoreans, and many others, down to the geometric logic of Ramon Lull (1232–1316), who was himself steeped in the *Kabbalah*. The variety of oracles that people have consulted for advice about the conduct of their lives—tarot cards, crystal balls, astrology, the entrails of animals and birds, palmistry, and the like—seems endless. For the purposes of this book, however, we shall be interested only in those aspects of divination that involve mathematics, such as magic squares. Whether or not a person believes that divination reveals hidden truth about the universe—the author does not—it remains a prominent form of human behavior over the centuries and deserves to be studied for that reason alone. But it is time to return to more prosaic matters.

The primary mathematical example of symbolism is the writing of positive integers. Some systems, like those of ancient Egypt, Greece, and Rome, are adequate for recording numbers, but comparatively cumbersome in computation. Just imagine trying to multiply XLI by CCCIV! [However, Detlefsen et al. (1975) demonstrate that this task is not as difficult as it might seem.] Even to use a 28×19 table of dates of Easter compiled in Russia some centuries ago, the calculators had to introduce simplifications to accommodate the fact that dividing a four-digit number by a two-digit number was beyond the skill of many of the users of the table.

The earliest mathematical texts discuss arithmetical operations using everyday words that were probably emptied of their usual meaning, thereby becoming abstract symbols capable of representing a variety of objects. Students had to learn to generalize from a particular example to the abstract case, and many problems that refer to specific objects probably became archetypes for completely abstract reasoning, just as we use such expressions as “putting the cart before the horse” and “comparing apples and oranges” to refer to situations having no connection at all with horse-and-buggy travel or the appraisal of fruit. For example, problems of the type “If 3 bananas cost 75 cents, how much do 7 bananas cost?” occur in the work of Brahmagupta from 1300 years ago. Brahmagupta named the three data numbers *argument* (3), *fruit* (75), and *requisition* (7). His rule for getting the answer was to multiply the fruit by the requisition and divide by the argument, a rule now known as the Rule of Three. As another example, cuneiform tablets from Mesopotamia that are several thousand years old contain general problems that we would now solve using quadratic equations. These problems are stated as variants of the problem of finding the length and width of a rectangle whose area and perimeter are known. The mathematician and historian of mathematics B. L. van der Waerden (1903–1996) claimed that the words for *length* and *width* were being used in a completely abstract sense in these problems. They had become abstract symbols, rather than words denoting concrete objects.

In algebra, symbolism seems to have occurred for the first time in the work of the (probably third- or fourth-century) Greek mathematician Diophantus of Alexandria, who introduced the symbol ζ for an unknown number. A document from India, the Bakshali manuscript, which may have been written within a century of the work of Diophantus, also introduces an abstract symbol for an unknown number. Symbolism developed gradually in modern algebra. Originally, the Arabic word for *thing* was used to represent the unknown in a problem. This word, and its Italian translation *cosa*, was eventually replaced by the familiar x most often used today. In this way an entire word was gradually pared down to a single letter that could be manipulated graphically.

2.4. MATHEMATICAL REASONING

Inferences made by animals and young children that appear to fit the “if A, then B” pattern are usually traceable to conditioning. When people or animals experience B after A, they rather quickly come to expect B any time that A happens, especially if the first experience had powerful emotional connections.

2.4.1. Animal Reasoning

Logic is concerned with getting conclusions that are as reliable as the premises. From a behavioral point of view, the human tendency to make inferences based on logic is probably hardwired and expressed as the same mechanism by which habits are formed. This same mechanism probably accounts for the metaphysical notion of *cause*. If A implies B , one feels that in some sense A *causes* B to be true. The dogs in Pavlov’s experiments, described above, were given *total* reinforcement as they learned geometry and came to make associations based on the constant conjunction of a given shape and a given reward or lack of reward. In the real world, however, we frequently encounter a weaker type of cause, where A is usually, but not always, followed by B . For example, lightning is always followed by thunder; but if the lightning is very distant, the thunder will not be heard. The analog of this weaker kind of cause in conditioning is *partial reinforcement*. A classical example is a famous experiment of Skinner (1948), who put hungry pigeons in a cage and attached a food hopper to the cage with an automatic timer to permit access to the food at regular intervals. The pigeons at first engaged in aimless activity when not being fed, but tended to repeat whatever activity they happened to be doing when the food arrived, as if they made an association between the activity and the arrival of food. Naturally, the more they repeated a given activity, the more likely that activity was to be reinforced by the arrival of food. Since they were always hungry, it was not long before they were engaged full time in an activity that they apparently considered an infallible food producer. This activity varied from one bird to another. One pigeon thrust its head into an upper corner of the cage; another made long sweeping movements with its head; another tossed its head back; yet another made pecking motions toward the floor of the cage.

The difficulties that people, even mathematicians, have in understanding and applying probability can be seen in this example. For example, the human body has some capacity to heal itself. Like the automatic timer that eventually provided food to the pigeons, the human immune system often overcomes the disease. Yet sick people, like hungry pigeons, try various methods of alleviating their misery. The consequence is a wide variety of nostrums said to cure a cold or arthritis. One of the triumphs of modern mathematical statistics is the

establishment of reliable systems of inference to replace the inferences that Skinner called “superstitious.”

Modern logic has purged the concept of implication of all connection with the notion of cause. The statement “If Abraham Lincoln was the first President of the United States, then $2 + 2 = 4$ ” is considered a true implication, even though Lincoln was not the first President and in any case his being such would have no *causal* connection with the truth of the statement “ $2 + 2 = 4$.” In standard logic the statement “If A is true, then B is true” is equivalent to the statement “Either B is true, or A is false, or both.” Absolute truth or falsehood is not available in relation to the observed world, however. As a result, science must deal with propositions of the form “If A is true, then B is *highly probable*.” One cannot infer from this statement that “If B is false, then A is highly *improbable*.” For example, an American citizen, taken at random, is probably not a U. S. Senator. It does not follow that if a person *is* a U. S. Senator, that person is probably not an American citizen.

Since we cannot trace the development of mathematical reasoning through the sources we have been using, we look instead at ancient documents and at the modern school curriculum to see how it arose. Taking the latter first, we note that students generally learn all of arithmetic and the rules for manipulating algebraic expressions by rote. Any justification of these rules is purely experimental. Logic enters the curriculum, along with proof, in the study of geometry. This sequence is not historical and may leave the impression that mathematics was an empirical science until the time of Euclid (ca. 300 BCE). But the ancient documents give us good reason to believe that some facts were *deduced* from simpler considerations at a very early stage. The main reason for thinking so is that the conclusions reached by some ancient authors are not visually obvious.

2.4.2. Visual Reasoning

As an example, it is immediately obvious that a diagonal divides a rectangle into two congruent triangles. If through any point on the diagonal we draw two lines parallel to the sides, these two lines will divide the rectangle into four rectangles. The diagonal divides two of these smaller rectangles into pairs of congruent triangles, just as it does the whole rectangle, thus yielding three pairs of congruent triangles, one large pair and two smaller pairs. It then follows (see Fig. 2.2) that the two remaining rectangles must have equal area, even though their shapes are different and *to the eye they do not appear to be equal*. Each of these rectangles is obtained by subtracting the two smaller triangles from the large triangle in which they are contained. When we find an ancient author mentioning that these two rectangles of different shape are equal, as if it were a well-known fact, we can be confident

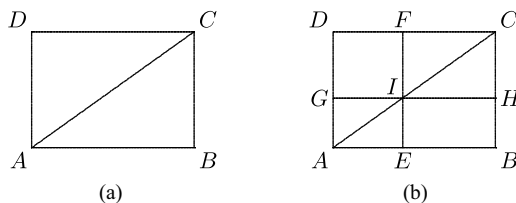


Figure 2.2. (a) The diagonal AC divides the rectangle $ABCD$ into congruent triangles ABC and CDA . (b) When the congruent pairs (AEI, IGA) and (IHC, CFI) are subtracted from the congruent pair (ABC, CDA) , the remainders (rectangles $EBHI$ and $GIFD$) must be equal.

that this knowledge does not rest on an experimental or inductive foundation. Rather, it is the result of a combination of numerical and spatial reasoning.

Ancient authors often state *what* they know without saying *how* they know it. As the example just cited shows, we can be confident that the basis was not always induction or experiment. Perminov (1997) points out that solutions of complicated geometric problems which can be shown to be correct are stated without proof by the writers of the very earliest mathematical documents, such as the Rhind papyrus from Egypt and cuneiform tablets from Mesopotamia. The facts that an author presents not merely a solution but a sequence of steps leading to that solution and that this solution can now be reconstructed justify the conclusion that the result was arrived at through mathematical reasoning, even though the author does not write out the details. This observation is particularly important in evaluating the mathematical achievements of the Mesopotamian, Egyptian, Hindu, and Chinese mathematicians, who did not write out formal proofs in the Greek sense. Since they often got results that agree with modern geometry, they must have used some form of visual reasoning such as we have presented here.

PROBLEMS AND QUESTIONS

Mathematical Problems

- 2.1. Find a unicursal tracing of the graph shown in Fig. 2.1.
- 2.2. Perminov (1997, p. 183) presents the following example of tacit mathematical reasoning from an early cuneiform tablet. Given a right triangle ACB divided into a smaller right triangle DEB and a trapezoid $ACED$ by the line DE parallel to the leg AC , such that EC has length 20, EB has length 30, and the trapezoid $ACED$ has area 320, what are the lengths AC and DE ? (See Fig. 2.3b.) The author of the tablet very confidently computes these lengths by the following sequence of operations: (1) $320 \div 20 = 16$; (2) $30 \cdot 2 = 60$; (3) $60 + 20 = 80$; (4) $320 \div 80 = 4$; (5) $16 + 4 = 20 = AC$; (6) $16 - 4 = 12 = DE$. As Perminov points out, to present

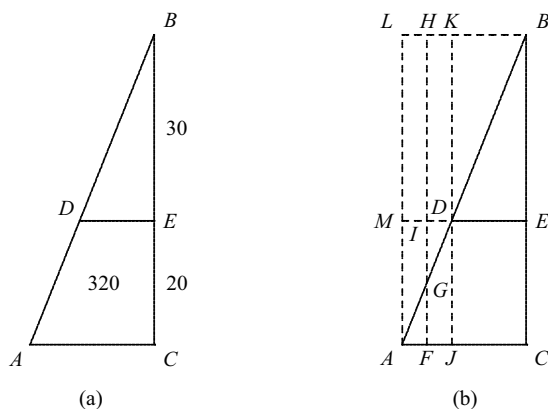


Figure 2.3. (a) Line DE divides triangle ABC into triangle DEB and trapezoid $ACED$. (b) Line $FGIH$ bisects line AD . Rectangle $FCEI$ has the same area as trapezoid $ACED$, and rectangle $JCED$ equals rectangle $MDKL$.

this computation with any confidence, you would have to know exactly what you are doing. What *was* this anonymous author doing?

To find out, fill in the reasoning in the following sketch. The author's first computation shows that a rectangle of height 20 and base 16 would have exactly the same area as the trapezoid. Hence if we draw the vertical line FH through the midpoint G of AD , and complete the resulting rectangles as in Fig. 2.3, rectangle $FCEI$ will have area 320. Since $AF = MI = FJ = DI$, it now suffices to find this common length, which we will call x ; for $AC = CF + FA = 16 + x$ and $DE = EI - DI = 16 - x$. By the principle demonstrated in Fig. 2.2, $JCED$ has the same area as $DKLM$, so that $DKLM + FJDI = DKLM + 20x$. Explain why $DKLM = 30 \cdot 2 \cdot x$, and hence why $320 = (30 \cdot 2 + 20) \cdot x$.

Could this procedure have been obtained experimentally?

- 2.3. A now-famous example of mathematical blunders committed by mathematicians (not statisticians, however) occurred a few decades ago. At the time, a very popular television show in the United States was called *Let's Make a Deal*. On that show, the contestant was often offered the chance to keep his or her current winnings or to trade them for a chance at some other unknown prize. In the case in question the contestant had chosen one of three boxes, knowing that only one of them contained a prize of any value, but not knowing the contents of any of them. For ease of exposition, let us call the boxes A, B, and C and assume that the contestant chose box A.

The emcee of the program was about to offer the contestant a chance to trade for another prize, but in order to make the program more interesting, he had box B opened, in order to show that it was empty. Keep in mind that the emcee *knew* where the prize was and would not have opened box B if the prize had been there. Just as the emcee was about to offer a new deal, the contestant asked to exchange the chosen box (A) for the unopened box (C) on stage. The problem posed to the reader is: Was this a good strategy? To decide, analyze 300 hypothetical games, in which the prize is in box A in 100 cases, in box B in 100 cases (in these cases, of course, the emcee will open box C to show that it is empty), and in box C in the other 100 cases. First assume that in all 300 games the contestant retains box A. Then assume that in all 300 games the contestant exchanges box A for the unopened box on stage. By which strategy does the contestant win more games?

- 2.4. Explain why the following analysis of the game described in the preceding question leads to an erroneous result. Consider all the situations in which the contestant has chosen box A and the emcee has shown box B to be empty. Imagine 100 games in which the prize is in box A and 100 games in which it is in box C. Suppose the contestant retains box A in all 200 games; then 100 will be won and 100 lost. Likewise, if the contestant switches to box C in all 200 games, then 100 will be won and 100 lost. Hence there is no advantage to switching boxes.
- 2.5. The fallacy discussed in the last two exercises is not in the mathematics, but rather in its application to the real world. The question involves what is known as *conditional probability*. Mathematically, the probability of event E, *given that event F has occurred*, is defined as the probability that E and F both occur, divided by the probability of F. The many mathematicians who analyzed the game erroneously proceeded by taking E as the event "The prize is in box C" and F as the event "Box B is empty." Given that box B has a $2/3$ probability of being empty and the event "E and F" is the same as

event E, which has a probability of $1/3$, one can then compute that the probability of E given F is $(1/3)/(2/3) = 1/2$. Hence the contestant seems to have a 50% probability of winning as soon as the emcee opens Box B, revealing it to be empty.

Surely this conclusion cannot be correct, since the contestant's probability of having chosen the box with the prize is only $1/3$ and the emcee can always open an empty box on stage. Replace event F with the more precise event "The emcee has *shown* that box B is empty" and redo the computation. Notice that the emcee is *going* to show that either box B or box C is empty and that the two outcomes are equally likely. Hence the probability of this new event F is $1/2$. Thus, even though the mathematics of conditional probability is quite simple, it can be a subtle problem to describe just what event has occurred. Conclusion: To reason correctly in cases of conditional probability, *one must be very clear in describing the event that has occurred.*

- 2.6. Reinforcing the conclusion of the preceding question, exhibit the fallacy in the following "proof" that *lotteries are all dishonest.*

Proof. The probability of winning a lottery is less than one chance in 1,000,000 ($= 10^{-6}$). Since all lottery drawings are independent of one another, the probability of winning a lottery five times is less than $(10^{-6})^5 = 10^{-30}$. But this probability is far smaller than the probability of any conceivable event. Any scientist would disbelieve a report that such an event had actually been observed to happen. Since the lottery has been won five times in the past year, it must be that winning it is not a random event; that is, the lottery is fixed.

What is the event that has to occur here? Is it "Person A (specified in advance) wins the lottery," or is it "At least one person in this population (of, say, 30 million people) wins the lottery"? What is the difference between those two probabilities? (The same fallacy occurs in the probabilistic arguments purporting to prove that evolution cannot occur, based on the rarity of mutations.)

Questions for Reflection

- 2.7. At what point do you find it necessary to count in order to say how large a collection is? Can you look at a word such as *tendentious* and see immediately how many letters it has? The American writer Henry Thoreau (1817–1863) was said to have the ability to pick up exactly one dozen pencils out of a pile. Try as an experiment to determine the largest number of pencils you can pick up out of a pile without counting. The point of this exercise is to see where direct perception needs to be replaced by counting.
- 2.8. How confidently can we make inferences about the development of mathematics from the study of animals, children, and archaeological sites?
- 2.9. One aspect of symbolism that has played a large role in human history is the mystical identification of things that exhibit analogous relations. The divination practiced by the Malagasy is one example, and there are hundreds of others: astrology, alchemy, numerology, tarot cards, palm reading, and the like, down to the many odd beliefs in the effects of different foods based on their color and shape. Even if we dismiss the validity of such divination, is there any value for science in the development of these subjects?

THE MIDDLE EAST, 2000–1500 BCE

In the five chapters that constitute this part of our study, we examine the mathematics produced in two contemporaneous civilizations, in Mesopotamia and Egypt, over a period from about 4000 to 3500 years ago. We shall look at the way each of these societies wrote numbers and calculated with them, and we shall discuss the uses they made of their calculations in geometry and applied problems.

Contents of Part II

1. Chapter 3 (Overview of Mesopotamian Mathematics) sketches the archaeological and mathematical background needed to appreciate the mathematical achievements recorded on the Old Babylonian tablets.
2. Chapter 4 (Computations in Ancient Mesopotamia) discusses some of the arithmetical and algebraic problems solved on the cuneiform tablets.
3. Chapter 5 (Geometry in Mesopotamia) looks at the area and volume problems solved by Mesopotamian mathematicians and their use of the Pythagorean theorem.
4. Chapter 6 (Egyptian Numerals and Arithmetic) introduces the numbering system used in ancient Egypt and the idiosyncratic method of multiplying by repeated doubling that is characteristic of this culture.
5. Chapter 7 (Algebra and Geometry in Ancient Egypt) discusses the applications of these numerical techniques made by the Egyptians in the areas of surveying, commerce, and engineering.

Overview of Mesopotamian Mathematics

Some quite sophisticated mathematics was developed four millennia ago in the portion of the Middle East that now forms the territory of Iraq and Turkey. This mathematics, along with a great deal of other lore, was written on small clay tablets in a style known as cuneiform (wedge-shaped), each tablet devoted to a limited topic. Nothing like a systematic treatise contemporary with this early mathematics exists. Scholars have had to piece together a mosaic picture of this mathematics from a few hundred clay tablets that show how to solve particular problems.

3.1. A SKETCH OF TWO MILLENNIA OF MESOPOTAMIAN HISTORY

The region known as Mesopotamia (Greek for “between the rivers”) was the home of many successive civilizations. The name of the region derives from the two rivers, the Euphrates and the Tigris, that flow from the mountainous regions around the Mediterranean, Black, and Caspian seas into the Persian Gulf. In ancient times this region was invaded and conquered many times, and the successive dynasties spoke and wrote in many different languages. The long-standing convention of referring to all the mathematical texts that come from this area as “Babylonian”—a term used as early as 450 BCE by the Greek historian Herodotus—gives undue credit to a single one of the many dynasties that dominated this region. Nevertheless, the appellation does fit the present discussion, since the tablets we are going to discuss are written in Old Babylonian.

Although many different peoples invaded this region over time, occupying different parts of it, we are going to oversimplify this history and divide it into eight different civilizations, as follows:

1. *Sumerian*. The Sumerians were either the original inhabitants of the region or immigrants from farther east. They spoke a language unrelated to the Semitic and Indo-European groups. They held sway over this region for several hundred years, starting about 3000 BCE. It was the Sumerians who invented the cuneiform writing, made by pressing a stylus into wet clay. Many of the small clay tablets containing such records dried out and have kept their information for over 4000 years.
2. *Akkadian*. These people were conquerors who spoke a Semitic language and adapted the Sumerian cuneiform writing to their own language. One consequence was the

compilation of Sumerian–Akkadian dictionaries, very useful for the later deciphering of these documents. The Akkadians established a commercial empire under King Sargon (ca. 2371–2316 BCE), which eventually collapsed and was replaced by a system of city–states in which the city of Ur at the mouth of the Euphrates was dominant.

3. *Amorite*. The Amorites, like the Akkadians, spoke a Semitic language. They invaded the area just before 2000 BCE and established a number of small kingdoms, of which Assyria was the first to become prominent, soon to be succeeded by Babylon under Hammurabi (1792–1750 BCE). This was the time when the Old Babylonian mathematical tablets were written. These tablets are the ones that will be discussed in the present chapter and the two following.
4. *Hittite*. The Hittites expanded from the west, the region now called Turkey. They spoke a language of the Indo-European family (the family to which English belongs). By 1650 BCE they had established a kingdom to rival the Amorites, and in 1595 they sacked the city of Babylon. The Hittite civilization collapsed around 1200 BCE due ultimately to pressure from the west exerted by the “Sea Peoples,” among whom were the Peleset, a people known to us from the Bible as the Philistines. They are the source of the name *Palestine*.
5. *Assyrian*. The Sea Peoples, although they caused the collapse of the Hittite Empire, did not occupy the portion of Mesopotamia that had been part of that empire. Instead, an empire based in the old city of Assyria began to grow and expand as far as its very well organized army and clever diplomacy could sustain it. The Assyrians eventually controlled a large portion of the region between the Mediterranean and the Persian Gulf, including present-day Palestine and parts of northern Egypt. Since this empire included the city of Babylon, it absorbed a great deal of the culture associated with that city. The Assyrian Empire was finally conquered by the Chaldean King Nebuchadnezzar (605–562 BCE).
6. *Chaldean*. This empire, although very short-lived (ca. 625–539 BCE), is well-known in the West because of Nebuchadnezzar, who is mentioned in the books of Kings, Jeremiah, and Daniel in the Bible. It was Nebuchadnezzar who conquered Jerusalem in 597 BCE and took the King of Judah and his followers into exile in Babylon. This civilization exerted a great influence on the writers of the Bible, especially the customs of the Chaldean court, where astrology was taken seriously.
7. *Persian*. As is well known from the Book of Daniel, the Chaldean empire was conquered in 539 BCE by the Persian king Cyrus the Great. Cyrus repatriated the exiles from Jerusalem and allowed the rebuilding of the Temple. The Persians, who speak an Indo-European language, have had an unbroken civilization since that time, although one subject to many changes of dynasty and religion. We shall see them coming into the story of mathematics at various points.
8. *Seleucid*. The high period of culture in mainland Greece coincided with the rise of the Athenian Empire in the middle of the fifth century BCE. The Athenian Empire was perceived as a threat by the Spartans, who brought it down through the Peloponnesian War (431–404 BCE). By that time, however, Greek scholarship and the Greek language were well established as intellectual forces. When the Macedonian kings Philip and Alexander conquered the territory eastward from mainland Greece to India and westward along the African coast of the Mediterranean, they consciously attempted to spread this culture. As a result, intellectual centers grew up in widely

separated places where scholars, not all Greek by birth, wrote and argued in the Greek language. The three best-known Greek mathematicians, Euclid, Archimedes, and Apollonius, lived and worked in Egypt, Sicily, and what is now Turkey.

When Alexander died in 323 BCE, his empire was divided among three of his generals. Besides the original Macedonian kingdom centered at Pella just north of Greece, there were two other regions with centers in Egypt and the Fertile Crescent. Egypt was ruled by the general Ptolemy Soter (the last of his heirs was Cleopatra, who presided over the incorporation of Egypt into the Roman Empire under Julius Caesar) while the regions around the Fertile Crescent were ruled by general Seleucus and thereby became known as the Seleucid Kingdom.

Unfortunately, in an introductory course, we cannot provide full details of the development of mathematics over this vast period of time. The reader should bear in mind that the mathematical examples in this chapter and the two following are a limited selection, wrenched out of their context. We are focusing on just a few salient features of one or two periods in this long and complicated history and are cherry-picking only the mathematics that seems most likely to interest the modern reader.

3.2. MATHEMATICAL CUNEIFORM TABLETS

Of the many thousands of cuneiform texts scattered through museums around the world, several hundred have been found to be mathematical in content. Deciphering them was made simpler by multilingual tablets that were created because the cuneiform writers themselves had need to know what had been written in earlier languages. A considerable amount of the credit for the decipherment must go to Sir Henry Rawlinson (1810–1895), who spent several years transcribing a trilingual inscription carved in a cliff at what is now Bisutun, Iran. This inscription, in Old Persian (an Indo-European language), Babylonian (a Semitic language derived from Akkadian), and Elamite (a “language isolate,” having no close relatives), tells of the reign of the Persian king Darius, who was successor to Cyrus the Great and reigned from 522 to 486. Its decipherment led to the recovery of the Akkadian language, which had gone extinct in the first century CE; and Akkadian led to the recovery of the Sumerian language, the language of the earliest civilization in Mesopotamia. Sumerian and Akkadian were freely mixed over a period of centuries and nearly melded into a single language.

By 1854, enough tablets had been deciphered to reveal the system of computation used in ancient Mesopotamia, and by the early twentieth century a considerable number of mathematical texts had been deciphered and analyzed. A detailed analysis of the ones known up to 1935 was presented in a two-volume work by Otto Neugebauer (1899–1992), *Mathematische Keilschrifttexte*, republished by Springer-Verlag in 1973. A more up-to-date study has been published by the Oxford scholar Eleanor Robson (1999).

Some of the tablets that have been discussed by historians of mathematics appear to be “classroom materials,” written by teachers as exercises for students. One clue that points toward this conclusion is that the answers so often “come out even.” As Robson (1995, p. 11, quoted by Melville, 2002, p. 2) states, “Problems were constructed from answers known beforehand.” See also the more recent book of Robson (2008, p. 21). Melville provides an example of a different kind from tablet YBC 4652 of the Yale Babylonian Collection in which the figures are not “rigged,” but a certain technique is presumed. Although there is an unavoidable lack of unity and continuity in the Mesopotamian texts compared with

mathematics written on media more amenable to extensive treatises, such as papyrus and paper, the cuneiform tablets nevertheless contain many problems similar to problems studied in other places such as India, China, and Egypt.

The applications that were made of these techniques must be conjectured, but we may confidently assume that they were the same everywhere: commerce, government administration, and religious rites, all of which call for counting and measuring objects on the earth and making mathematical observations of the sky in order to keep track of months and years. According to Robson (2009, pp. 217–218), the education of a surveyor, which required both numerical and geometric skill, was of crucial importance in keeping public order:

When I go to divide a plot, I can divide it; when I go to apportion a field, I can apportion the pieces, so that when wronged men have a quarrel I soothe their hearts and [. . .]. Brother will be at peace with brother.

3.3. SYSTEMS OF MEASURING AND COUNTING

The systems of numeration still used in the United States, the last bastion of resistance to the metric system, show that people once counted by twos, threes, fours, sixes, eights, and twelves. In the United States, eggs and pencils, for example, are sold by the *dozen* or the *gross*. Until recently, stock averages were quoted in eighths rather than tenths. Measures of length, area, and weight show other groupings. Consider the following words: *fathom* (6 feet), *foot* (12 inches), *pound* (16 ounces), *yard* (3 feet), *league* (3 miles), *furlong* (1/8 of a mile), *dram* (1/8 or 1/16 of an ounce, depending on the context), *karat* (1/24, used as a pure number to indicate the proportion of gold in an alloy),¹ *peck* (1/4 of a bushel), *gallon* (1/2 peck), *pint* (1/8 of a gallon), and *teaspoon* (1/3 of a tablespoon). The strangest unit of all in the formidable English system—no longer the *English* system since the UK became part of the European Union—is the acre, 1/640 of a square mile. In the United States, a square mile was called a *section*, and farms commonly consisted of a quarter of a section, 160 acres. (In metric units, an acre is about 0.4 hectares.)

Even in science, however, there remain some vestiges of nondecimal systems of measurement inherited from the ancient Middle East. In the measurement of both angles and time, minutes and seconds represent successive divisions by 60. A day is divided into 24 hours, each of which is divided into 60 minutes, each of which is divided into 60 seconds. At that point, our division of time becomes decimal; we measure races in tenths and hundredths of a second. A similar renunciation of consistency came in the measurement of angles as soon as hand-held calculators became available. Before these calculators came into use, students (including the present author) were forced to learn how to interpolate trigonometric tables in minutes (1/60 of a degree) and seconds (1/60 of a minute). In physical measurements, as opposed to mathematical theory, we still divide circles into 360 equal degrees. But our hand-held calculators have banished minutes and seconds. They divide degrees decimally and of course make interpolation an obsolete skill. Since π is irrational, it seems foolish to adhere to any rational fraction of a circle as a standard unit; hand-held calculators are perfectly content to use the natural (radian) measure, and we could eliminate a useless button

¹The word is a variant of *carat*, which also means 200 milligrams when applied to the size of a diamond.

by abandoning degrees entirely. That reform, however, is likely to require even more time than the adoption of the metric system.²

3.3.1. Counting

A nondecimal counting system reported (1937) by the American mathematicians David Eugene Smith (1860–1944) and Jekuthiel Ginsburg (1889–1957) as having been used by the Andaman of Australia illustrates how one can count up to certain limits in a purely binary system. The counting up to 10, translated into English, goes as follows: “One two, another one two, another one two, another one two, another one two. That’s all.” In saying this last phrase, the speaker would bring the two hands together. This binary counting *appears* to be very inefficient from a human point of view, but it is the system that underlies the functioning of computers, since a switch has only two positions. The binary digits or *bits*, a term that seems to be due to the American mathematician Claude Shannon, are generally grouped into larger sets for processing.

Although bases smaller than 10 are used for various purposes, some societies have used larger bases. Even in English, the word *score* for 20 (known to most Americans only from the first sentence of Lincoln’s Gettysburg Address) does occur. In French, counting between 60 and 100 is by 20s. Thus, 78 is *soixante dix-huit* (sixty-eighteen) and 97 is *quatre-vingt dix-sept* (four-twenty seventeen). Menninger (1969, pp. 69–70) describes a purely *vigesimal* (base 20) system used by the Ainu of Sakhalin. Underlying this system is a base 5 system and a base 10 system. Counting begins with *shi-ne* (*begin-to-be* = 1), and progresses through such numbers as *aschick-ne* (*hand* = 5), *shine-pesan* (*one away from* [10] = 9), *wan* (*both sides* = *both hands* = 10), to *hot-ne* (*whole-* [person]-*to-be* = 20). In this system 100 is *ashikne hotne* or 5 twenties; 1000, the largest number used is *ashikne shine wan hotne* or 5 ten-twenties. There are no special words for 30, 50, 70, or 90, which are expressed in terms of the basic 20-unit. For example, 90 is *wan e ashikne hotne* (10 from 5 twenties). Counting by subtraction probably seems novel to most people, but it does occur in Roman numerals (*IV* = 5 – 1), and we use subtraction to tell time in expressions such as *ten minutes to four* and *quarter to five*.³

3.4. THE MESOPOTAMIAN NUMBERING SYSTEM

As the examples of angle and time measurement show, the successive divisions or regroupings in a number system need not have the same number of elements at every stage. The Mesopotamian sexagesimal system appears to have been superimposed on a decimal system. In the cuneiform tablets in which these numbers are written the numbers 1 through

²By abandoning another now-obsolete system—the Briggsian logarithms—we could eliminate *two* buttons on the calculators. The base 10 was useful in logarithms only because it allowed the tables to omit the integer part of the logarithm. Since no one uses tables of logarithms any more, and the calculators don’t care how messy a computation is, there is really no reason to do logarithms in any base except the natural one, the number *e*, or perhaps base 2 (in number theory). Again, don’t expect this reform to be achieved in the near future.

³Technology, however, is rapidly removing this last vestige of the old way of counting from everyday life. Circular clock faces have been largely replaced by linear digital displays, and ten minutes to four has become 3:50. This process began long ago when railroads first imposed standard time in place of mean solar time and brought about the first 24-hour clocks.

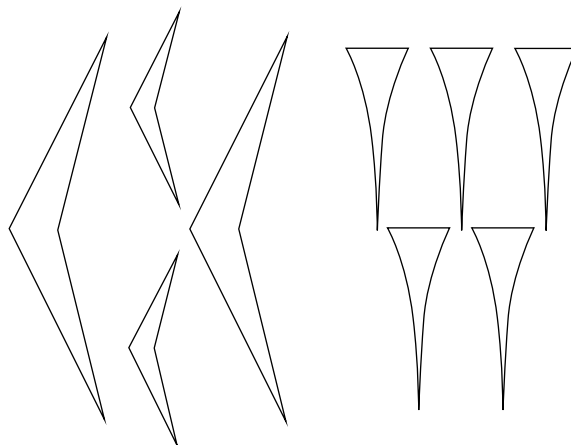


Figure 3.1. The cuneiform number 45.

9 are represented by a corresponding number of wedge-shaped vertical strokes, and 10 is represented by a new symbol, a hook-shaped mark that resembles a boomerang (Fig. 3.1). However, the next grouping is not *ten* groups of 10, but rather *six* groups of 10. Even more strikingly, the symbol for the next higher group is again a vertical stroke. Logically, this system is equivalent to a base-60 place-value system with a floating “decimal” (sexagesimal) point that the reader or writer had to keep track of mentally. Within each unit (sexagesimal rank) of this system there is a truncated decimal system that is not place-value, since the ones and tens are distinguished by different symbols rather than physical location.

3.4.1. Place-Value Systems

Since we take our familiar place-value decimal system for granted, it is worth remembering that several advanced civilizations, including those of Egypt, ancient Greece, and ancient Rome, did not have such a system. The Egyptians and ancient Greeks (who probably copied the Egyptians in this matter) had, as we do, individual symbols for the numbers 1 through 9, but they had nine more symbols for 10 through 90 and another nine symbols for 100 through 900. (The Greeks used their 24-letter alphabet, along with three obsolete letters, to get these 27 symbols.) Even the Chinese system, which was decimal, used separate symbols for each power of 10. For example, the numbers 1, 2, and 3 are symbolized as $-$, $=$, and \equiv , while 10 was represented by a cross shape ($+$ or \dagger). But 20 was written as $=+$, that is, a 2 whose value was shown by being attached to the symbol for the corresponding power of 10. This extra symbol, we can now see, was not needed if you have a symbol for an empty place, since the physical location of the 2 suffices to show its value. Thus, the Chinese system was “just short of” a full place-value system.

It is therefore somewhat surprising that a pure place-value sexagesimal system arose as early as 4000 years ago in Mesopotamia, with which Egypt, Greece, and Rome were in contact. Somehow, the advantages of the system penetrated only Greco-Roman science, not commerce and other economic activity.⁴ In its original form, this system lacked one

⁴Perhaps, considering the cries of outrage whenever any attempt is made to use the metric system in the United States, we should not be surprised.

feature that we regard as essential today, a symbol for an empty place (zero). The later Greek writers, such as Ptolemy in the second century CE, used the sexagesimal notation with a circle to denote an empty place.

Although it obviously began as a decimal system, since there are distinct symbols for 1 and 10, the special symbol for 100, which one would expect in such a system, does not occur in the clay tablets from which most of our knowledge about Mesopotamian mathematics is derived. The reason is that at some point the Mesopotamians developed a true place-value system of writing numbers, using 60 as a base. This system was taken over, but for scientific purposes only, by the Greeks, who passed on to the modern world the idea of dividing a day into 24 hours, an hour into 60 minutes, a minute into 60 seconds, and a circle into 360 degrees. To get a picture of the way the writers were thinking that doesn't require the use of non-standard symbols, historians of the subject have invented a way of transcribing the numbers into easily recognizable forms. We shall now describe this transcription.

3.4.2. The Sexagesimal Place-Value System

You are familiar with the fact that the number 3926 means $3 \times 10^3 + 9 \times 10^2 + 2 \times 10^1 + 6 \times 10^0$, that is, $3000 + 900 + 20 + 6$. We are working in base 10 here, and each digit will be an integer between 0 and 9.

If we were interpreting it in base 60, the symbol 3926 would mean $3 \times 60^3 + 9 \times 60^2 + 2 \times 60 + 6$, that is, $648000 + 32400 + 120 + 6$, or 680,526 in decimal notation. We could write this equality as $3926_{60} = 680526_{10}$. However, we shall normally omit the subscripts, since it will be obvious which of the two bases is meant.

In sexagesimal notation, each digit is between 0 and 59, and that fact gives rise to some ambiguity: How can we be sure the number we just looked at was not meant to be, for example, $39 \times 60 + 26$? The distinction would be clear in authentic cuneiform notation, since there is a special symbol for 10. To make it clear in our transcription, we will use a comma to separate the digits of all sexagesimal numbers. In that way, we can distinguish between 3, 9, 2, 6 and 39, 26. Between the integer and fractional parts of the number, we shall write a semicolon in our transcription. Thus 35, 6; 12, 9 means $35 \times 60 + 6 + \frac{12}{60} + \frac{9}{60^2}$, which in decimal notation would be 2106.2025. As another example, the number that we write as 85.25 could be transcribed into this notation as 1, 25; 15, meaning $1 \cdot 60 + 25 \cdot 1 + 15 \cdot \frac{1}{60}$.

Converting a number written in sexagesimal notation into decimal notation is a very easy matter. Just insert the appropriate power of 60 (positive for digits left of the units digit, negative for digits right of it) in each place and multiply by the digit in that place; then add the products. This skill requires hardly any practice. You should be able to verify easily that 13,7;21 converts to $13 \times 60 + 7 + \frac{21}{60} = 787.35$ and that 2,29;15,11 converts to $3629 \frac{911}{3600} = 3629.25305555 \dots$, where the 5's repeat forever. As you see, a terminating sexagesimal number may fail to terminate when translated into decimal notation. A terminating decimal number, however, will always terminate when converted to sexagesimal notation, because 10 divides 60. Our first task is to spend a little time converting between the hybrid notation just introduced for the sexagesimal system and the decimal system we are familiar with, so that numbers written in this system will appear less strange.

3.4.3. Converting a Decimal Number to Sexagesimal

The procedure for converting from decimal notation to sexagesimal requires separate handling of the integer and fractional parts of a number. We begin by discussing how to convert

integers. The general procedure is sufficiently well illustrated by the conversion of the decimal number 3,874,065 into sexagesimal notation. The rule is to do repeated long division with 60 as the divisor, using the remainder at each stage as the sexagesimal digit and the quotient as the next dividend, until finally a quotient less than 60 is obtained and used as the leftmost digit:

$$3874065 \div 60 = 64567$$

with a remainder of 45. Hence the units place is 45. Now we divide the quotient (64567) by 60:

$$64567 \div 60 = 1076$$

with a remainder of 7. Hence the 60-digit is 7. We then divide 1076 by 60:

$$1076 \div 60 = 17$$

with a remainder of 56, so that the 60^2 -digit is 56, and now the 60^3 -digit is immediately seen to be 17. Thus

$$3,874,065_{10} = 17, 56, 7, 45_{60}.$$

All you have to remember is that you are working *leftward* from the “sexagesimal point” (the semicolon). You can verify that this is correct by converting in the opposite direction: $17 \times 60^3 + 56 \times 60^2 + 7 \times 60 + 45 = 3,874,065$.

We next show how to convert a fractional number from either common-fraction or decimal-fraction form into sexagesimal form. Since multiplying a number by 60 is equivalent to moving the sexagesimal point to the right, the basic principle is that these successive sexagesimal digits reveal themselves as the integer part of the product when the number is repeatedly multiplied by 60. This principle is completely obvious and trivial if the fraction is given in sexagesimal form to begin with. For example, suppose the number is $N = 0; 43, 12, 19$. Then $60N = 43; 12, 19$, so that the first digit of the original fractional number N is the integer part (43) of $60N$. After discarding that integer part, we can get the second digit by multiplying what is left again by 60, and obviously that will be 12 in the present case. This procedure works whether or not the fraction is given in sexagesimal form. All we have to remember is to carry out the procedure “dual” to the procedure just described for converting integers. In this dual procedure, you repeatedly *multiply* by 60 and take the *integer parts* of the products as the successive digits. This time, you are working rightward, again away from the “sexagesimal point.”

We illustrate with the fraction $\frac{2}{25}$. We find

$$\frac{2 \times 60}{25} = \frac{120}{25} = 4 \frac{20}{25} = 4 \frac{4}{5}.$$

Thus, the first digit right of the sexagesimal point is 4. To get the second one, we need to convert the fraction $\frac{4}{5}$, which is done exactly the same way:

$$\frac{4 \times 60}{5} = 48.$$

Thus, we have found that

$$\frac{2}{25} = 0; 4, 48.$$

And you can verify that this is correct:

$$\frac{4}{60} + \frac{48}{60^2} = \frac{1}{15} + \frac{1}{75} = \frac{6}{75} = \frac{2}{25}.$$

Decimal fractions can be converted by following this procedure, using a hand calculator to facilitate the computation, or by changing it to a common fraction. For example, we could convert 0.337 to $\frac{337}{1000}$ and proceed:

$$\frac{60 \times 337}{1000} = \frac{3 \times 337}{50} = \frac{1011}{50} = 20 \frac{11}{50}.$$

The first digit is thus 20. To get the next one we continue:

$$\frac{60 \times 11}{50} = \frac{66}{5} = 13 \frac{1}{5}.$$

Thus second digit is now seen to be 13, and we get the third and final digit by converting $\frac{1}{5}$ to $\frac{12}{60}$. Hence

$$0.337_{10} = 0; 20, 13, 12_{60}.$$

Peculiarities to Watch For. If you have a repeating decimal for which you know an exact value as a common fraction, for example, $0.33333 \dots$, which you know is $\frac{1}{3}$, it is best to convert it to the common fraction before converting it to sexagesimal. The procedure given above will correctly convert $\frac{1}{3}$ into $0; 20$. But if you work with its infinite decimal expression $0.3333 \dots$, you will first of all have difficulty multiplying it by 60, since the multiplication has to start at the right-hand end, which is infinitely distant. Even if you do the obvious thing and say that $0.333 \dots \times 60 = 19.99999 \dots$, you will get 19 as the first digit and then have to convert $0.99999 \dots$, which will similarly yield $59.999999 \dots$ when multiplied by 60. Hence you'd find that $0.333 \dots_{10} = 0; 19, 59, 59, 59, \dots_{60}$, which is correct, but clumsily expressed.

Some fractions do have nonterminating sexagesimal expansions. For example, $\frac{2}{7}$ will repeat with period 3:

$$\frac{2}{7} = 0; 17, 8, 34, 17, 8, 34, \dots$$

3.4.4. Irrational Square Roots

Some of the tablets contain the sexagesimal number $1;24,51,10$, which is $\frac{30547}{21600} \approx 1.41421296296\dots$. It is clear from the context that this number is being used as an approximation to $\sqrt{2}$. The tablet YBC 7289 in the Yale Babylonian Collection exhibits the number $0;42,30$ in a context where it clearly means $\frac{1}{\sqrt{2}}$. There are various ways in which these approximations might have been arrived at. The simplest conjecture amounts essentially to what eventually became generalized as the *Newton–Raphson* method of approximating the root of a functional equation $f(x) = 0$. In the limited context of the equation $x^2 - 2 = 0$, the method has a much simpler explanation than it gets in calculus courses.

We begin by noting that $\sqrt{2}$ must be between 1 and 2. Hence, let us begin with $\frac{3}{2}$ as an approximation. This number happens to be too large, but we do not have to know that in order to improve the approximation. If an approximation a is too large, then $\frac{2}{a}$ will be too small, and hence we are likely to improve the approximation by averaging a and $\frac{2}{a}$:

$$a \rightarrow \frac{1}{2} \left(a + \frac{2}{a} \right) = \frac{a}{2} + \frac{1}{a}.$$

Starting with $a = \frac{3}{2}$, we find the next approximation to be

$$\frac{3}{4} + \frac{2}{3} = \frac{17}{12}.$$

This is our new a , and we continue from there. If $\frac{17}{12}$ is taken as an approximation of $\sqrt{2}$, then $\frac{17}{24}$ is an approximation for $\frac{1}{\sqrt{2}}$, and that is the approximation actually used in the tablet YBC 7289. The next approximation to $\sqrt{2}$ is $\frac{577}{408} = 1 + \frac{169}{408}$, and its sexagesimal expansion begins $1;24,51,10$, which is the approximation used in the tablets.

This explanation is only a conjecture; we don't really know how square roots were found. In the next chapter, we shall discuss how computations were done within the sexagesimal system. You can imagine that it was not so easy as it is with our base-10 system.

PROBLEMS AND QUESTIONS

Mathematical Problems

- 3.1. Convert the sexagesimal number $11, 4, 29; 58, 7$ into decimal notation.
- 3.2. Convert the decimal number 4752.73 into sexagesimal notation.
- 3.3. Convert the improper fraction $\frac{9437}{2755}$ into sexagesimal notation. (Get the first four digits on the right of the sexagesimal point. This sexagesimal expansion does eventually repeat. However, it has a powerfully long period!)

Historical Questions

- 3.4. What special circumstances made it possible to decipher the cuneiform tablets?
- 3.5. Why was it important for a government in ancient times to have a cadre of competent surveyors?

- 3.6. What are the differences in the counting systems used in ancient China, Egypt, and Mesopotamia?

Questions for Reflection

- 3.7. Why would it be more difficult to use a sexagesimal place-value system than a decimal place-value system? How would you overcome this difficulty if you had only the sexagesimal system to use?
- 3.8. What other bases, besides 10 and 60, have you heard of being used? Suppose one person was using base 7 and another base 12. What advantages would each have over the other?
- 3.9. How might a sexagesimal system have originated in the first place, since by far the commonest bases used throughout the world are 10, 5, and 20?

Computations in Ancient Mesopotamia

We might expect that, with their place-value system, the ancient Mesopotamians would have done arithmetic somewhat as we do. There are, however, a few differences. While we have no need to discuss addition and subtraction, we do need to compare multiplication and division in the two systems.

4.1. ARITHMETIC

Cuneiform tablets at the British Museum from the site of Senkereh (also known as Larsa) contain tables of products, reciprocals, squares, cubes, square roots, and cube roots of integers. It appears that the people who worked with mathematics in Mesopotamia learned by heart, just as we do, the products of all the small integers. Of course, for them a theoretical multiplication table would have to go as far as 59×59 , and the consequent strain on memory would be large. That fact may account for the existence of so many written tables. Just as most of us learn, without being required to do so, that $\frac{1}{3} = 0.3333 \dots$, the Mesopotamians wrote their fractions as sexagesimal fractions and probably came to recognize certain reciprocals, for example $\frac{1}{9} = 0; 6, 40$. With a system based on 30 or 60, all numbers less than 10 except 7 have terminating reciprocals. In order to get a terminating reciprocal for 7, one would have to go to a system based on 210, which would be far too complicated.

Even with base 60, multiplication can be quite cumbersome, and historians have conjectured that calculating devices such as an abacus might have been used, although none have been found. Høyrup (2002) has analyzed the situation by considering the errors in two problems on Old Babylonian cuneiform tablets and deduced that any such device would have had to be some kind of counting board, in which terms that were added could not be identified and subtracted again (like pebbles added to a pile).

If we try to reconstruct a base-60 multiplication using what we know of decimal multiplication, where we “carry” the tens digit to the next column to the left when the product of two digits is larger than 10, we find that it is possible to do so. However, any reasonably complicated multiplication will task the calculator’s patience, since the carrying involves a sixties digit rather than a tens digit. Here is a sample multiplication that, when translated into decimal notation, verifies the computation

$$\frac{84,387,829}{1200} \times \frac{4636}{15} = \frac{97,805,493,811}{4500}.$$

As you see, even without a calculator, the sexagesimal computation is only a little harder than the decimal calculation, especially since neither denominator has a terminating decimal expansion.

19,	32,	3;	11,	27	
	5,	9;	4		
	1,	18,	8;	12,	45, 48
	2,	55,	48,	28;	43, 3
1,	37,	40,	15,	57;	15
1,	40,	37,	22,	34;	10, 48, 48

4.1.1. Square Roots

Not only are sexagesimal fractions handled easily in all the tablets, the concept of a square root occurs explicitly, and actual square roots are approximated by sexagesimal fractions, showing that the mathematicians of the time realized that they hadn't been able to make these square roots come out even. Whether they realized that the square root would never come out even is not clear. For example, text AO 6484 (the AO stands for Antiquités Orientales) from the Louvre in Paris contains the following problem on lines 19 and 20:

The diagonal of a square is 10 Ells. How long is the side? [To find the answer] multiply 10 by 0;42,30. [The result is] 7;5.

Now $0;42,30$ is $\frac{42}{60} + \frac{30}{3600} = \frac{17}{24} \approx 0.7083$. This is the same approximation to $1/\sqrt{2} \approx 0.7071$ that is found on the tablet YBC 7289, discussed in the preceding chapter. The answer $7;5$ is $7\frac{1}{12} \approx 7.083 = 10 \cdot 0.7083$. It seems that the writer of this tablet knew that the ratio of the side of a square to its diagonal is approximately $\frac{17}{24}$. As mentioned in the preceding chapter, the approximation to $\sqrt{2}$ that arises from what is now called the *Newton–Raphson* method, starting from $\frac{3}{2}$ as the first approximation, turns up the number $\frac{17}{12}$ as the next approximation, and hence $\frac{17}{24}$ represents an approximation to $\frac{\sqrt{2}}{2} = \frac{1}{\sqrt{2}}$.

The writers of these tablets realized that when numbers are combined by arithmetic operations, it may be of interest to know how to recover the original data from the result. This realization is the first step toward attacking the problem of inverting binary operations. Although we now handle such problems by solving quadratic equations, the Mesopotamian approach did not involve any explicit mention of equations. Instead, many of the tablets show a routine procedure, associating with a pair of numbers, say 13 and 27, two other numbers: their average $(13 + 27)/2 = 20$ and their *semidifference*¹ $(27 - 13)/2 = 7$. The average

¹This word is coined because English contains no one-word description of this concept, which must otherwise be described as half of the difference of the two numbers. It is clear from the way in which the semidifference occurs constantly that the writers of these tablets automatically looked at this number along with the average when given two numbers as data. There seems to be no word in the Akkadian, Sumerian, and ideogram glossary given by Neugebauer to indicate that the writers of the clay tablets had a special word for these concepts. It seems clear, however, that the scribes were trained to calculate these numbers when dealing with this type of problem. In the translations given by Neugebauer, the average and semidifference are obtained one step at a time, by first adding or subtracting the two numbers and then taking half of the result.

and semidifference can be calculated from the two numbers, and the original data can be calculated from the average and semidifference. The larger number (27) is the sum of the average and semidifference: $20 + 7 = 27$, and the smaller number (13) is their difference: $20 - 7 = 13$. The realization of this mutual connection makes it possible essentially to “change coordinates” from the number pair (a, b) to the pair $((a + b)/2, (a - b)/2)$.

At some point lost to history, some Mesopotamian mathematician came to realize that the product of two numbers is the difference of the squares of the average and semidifference: $27 \cdot 13 = (20)^2 - 7^2 = 351$ (or 5, 51 in Mesopotamian notation). This principle made it possible to recover two numbers when knowing their sum and product or knowing their difference and product. For example, given that the sum is 10 and the product is 21, we know that the average is 5 (half of the sum), hence that the square of the semidifference is $5^2 - 21 = 4$. Therefore, the semidifference is 2, and the two numbers are $5 + 2 = 7$ and $5 - 2 = 3$. Similarly, knowing that the difference is 9 and the product is 52, we conclude that the semidifference is 4.5 and the square of the average is $52 + (4.5)^2 = 72.25$. Hence the average is $\sqrt{72.25} = 8.5$. Therefore, the two numbers are $8.5 + 4.5 = 13$ and $8.5 - 4.5 = 4$. The two techniques just illustrated occur constantly in the cuneiform texts and seem to be procedures familiar to everyone, requiring no explanation. At this point, the development of computational procedures has led to algebra, in the sense that the problems require turning an *implicit* definition of a number into an *explicit* numerical value.

The important principle here, that the difference of the squares of the average and semidifference is the product, was to have important consequences over the next four thousand years of mathematical progress, after it was combined with the Pythagorean theorem. The principle that was in the minds of the Mesopotamian mathematicians was a two-part procedure: (1) If you are given two numbers a and b , the numbers $c = (a + b)/2$ (their average) and $d = (a - b)/2$ (their semidifference) reveal important information about them; (2) the difference of the squares of c and d is the product of a and b . In the cuneiform tablets, this principle finds algebraic application, making it possible by taking the square root to find either the average or the semidifference, provided that you know the other and that you also know their product. When combined with the Pythagorean theorem, which the Mesopotamians also knew, this “polarization principle” can lead to even more interesting new results, and did so for nearly four thousand years.

4.2. ALGEBRA

If we interpret Mesopotamian algebra in our own terms, we can credit the mathematicians of that culture with knowing how to solve some systems of two linear equations in two unknowns, any quadratic equation having at least one real positive root, some systems of two equations where one of the equations is linear and the other quadratic, and a potentially complete set of cubic equations. Of course, it must be remembered that these people were solving *problems*, not *equations*. They did not have any classification of equations in which some forms were solvable and others not. What they knew was that they could find certain numbers from certain data. For that reason, the reader is cautioned to read the following subsection headings with reservations. The ancient mathematicians were solving problems that *we now solve* using algebra and classify according to these headings. They themselves must have had some other classification, since the concept of an equation did not yet exist.

4.2.1. Linear and Quadratic Problems

As just mentioned, the Mesopotamian approach to algebraic problems was to associate with every pair of numbers another pair: their average and their *semidifference*. These associations provide what we now call linear changes of variable. Linear problems arise frequently as a subroutine in the solution of more complex problems involving squares and products of unknowns. In Mesopotamia, what we now think of as quadratic equations occur most often as problems in two unknown quantities, usually the length and width of a rectangle. The Mesopotamian mathematicians were able to reduce a large number of problems to the form in which the sum and product or the difference and product of two unknown numbers are given. We shall consider an example that has been written about by many authors. It occurs on a tablet from the Louvre in Paris, known as AO 8862.

A loose translation of the text of this tablet, made from Neugebauer's German translation, reads as follows:

I have multiplied the length and width so as to make the area. Then I added to the area the amount by which the length exceeds the width, obtaining 3,3. Then I added the length and width together, obtaining 27. What are the length, width, and area?

27	3,3	the sums
15		length
3,0		area
12		width

You proceed as follows:

Add the sum (27) of the length and width to 3,3. You thereby obtain 3,30. Next add 2 to 27, getting 29. You then divide 29 in half, getting 14;30. The square of 14;30 is 3,30;15. You subtract 3,30 from 3,30;15, leaving the difference of 0;15. The square root of 0;15 is 0;30. Adding 0;30 to the original 14;30 gives 15, which is the length. Subtracting 0;30 from 14;30 gives 14 as width. You then subtract 2, which was added to the 27, from 14, giving 12 as the final width.

The author continues, verifying that these numbers do indeed solve the problem. This text requires some commentary, since it is baffling at first. Knowing the general approach of the Mesopotamian mathematicians to problems of this sort, one can understand the reason for dividing 29 in half (so as to get the average of two numbers) and the reason for subtracting 3,30 from the square of 14;30 (the difference between the square of the average and the product will be the square of the semidifference of the two numbers whose sum is 29 and whose product is 3,30, that is, 210). What is not clear is the following: Why add 27 to the number 3,3 in the first place, and why add 2 to 27? Possibly the answer is contained in Fig. 4.1, which shows that adding the difference between length and width to the area amounts to gluing a smaller rectangle of unit width onto the rectangle whose dimensions are to be found. Then adding the sum of length and width amounts to gluing a gnomon onto the resulting figure in order to complete a rectangle two units wider than the original. Finding the dimensions of that rectangle from its perimeter and area is the standard technique of solving a quadratic equation, and that is what the author does. It is not clear that this gluing of additions onto the rectangle represents the thought process of the original author. The present author finds this very plausible, but it is worth noting that van der Waerden

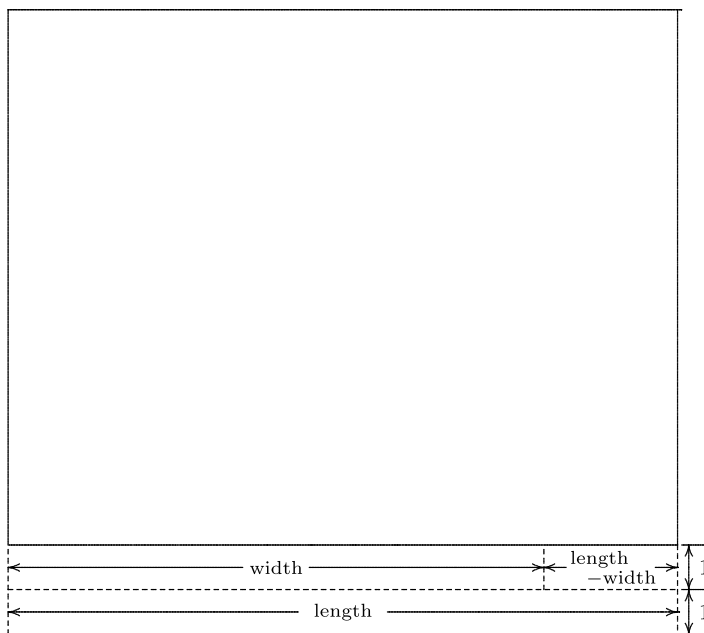


Figure 4.1. Reduction of a problem to standard form.

(1963) insisted that the original author was actually carrying out the mathematically absurd operation of adding length to area. On those grounds, he concluded that “we may safely set this down as a pair of equations in two unknowns.”²

The tablet AO 6670, as explicated by van der Waerden (1963, pp. 73–74), involves two unknowns and two conditions, given in abstract terms without specific numbers. Unfortunately, the explanation is very difficult to understand. The statement of the problem is taken directly from Neugebauer’s translation: *Length and width as much as area; let them be equal*. Thereafter, the translation given by van der Waerden, due to François Thureau-Dangin (1872–1944), goes as follows:

The product you take twice. From this you subtract 1. You form the reciprocal. With the product that you have taken you multiply, and the width it gives you.

Van der Waerden asserts that the formula $y = (1/(x - 1)) \cdot x$ is “stated in the text” of Thureau-Dangin’s translation. If so, it must have been stated in a place not quoted by van der Waerden, since x is not a “product” here, nor is it taken twice. Van der Waerden also notes that according to Evert Marie Bruins (1909–1990), the phrase “length and width” does not mean the *sum* of length and width. Van der Waerden says that “the meaning of the words has to be determined in relation to the mathematical content.” The last two sentences

²Van der Waerden also argued that the words for length and width (*uš* and *šag*), being indeclinable, were being used as symbols for an abstract unknown quantity. The choice is stark: We must conclude that the scribe either was using a kind of linguistic shorthand in which a length becomes a rectangle of unit width or was a modern algebraist for whom dimensional consistency is of no importance.

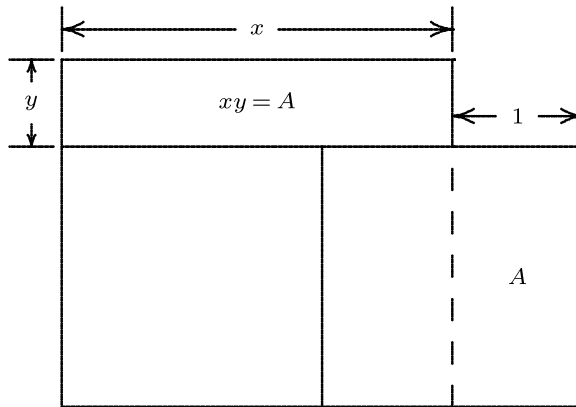


Figure 4.2. A scenario that may “fit” a text from cuneiform tablet AO 6670.

in the description tell how to determine the width once the length has been found. That is, you take the reciprocal of the length and multiply it by the product of length and width, which must be given in the problem as the area. The mystery is then pushed into the first two instructions. What product is being “taken twice”? Does taking a product twice mean multiplying by 2, or does it mean cubing? Why is the number 1 being subtracted? Perhaps we should go back to the original statement and ask whether “as much as area” implies an equation, or whether it simply means that length and width *form* an area. What does the word *them* refer to in the statement, “Let them be equal”? Is it the length and the width, or some combination of them and the area? Without knowing the original language and seeing the original text, we cannot do anything except suggest possible meanings, based on what is mathematically correct, to those who do know the language.

We can get a geometric problem that fits this description by considering Fig. 4.2, where two equal squares have been placed side by side and a rectangle of unit length, shown by the dashed line, has been removed from the end. If the problem is to construct a rectangle on the remaining base equal to the part that was cut off, we have conditions that satisfy the instructions in the problem. That is, the length x of the base of the new rectangle is obtained numerically by subtracting 1 from twice the given area. Still, it is dangerous for any nonspecialist to speculate about the meaning without being able to read the original document, and it is probably best to leave this problem at this point.

4.2.2. Higher-Degree Problems

Cuneiform tablets have been found—one such being VAT 8402, for example³ [see Neugebauer (1935, p. 76)]—that give the sum of the square and cube of an integer for many values of the integer. These tablets may have been used for finding the numbers to which this operation was applied in order to obtain a given number. In modern terms these tablets make it possible to solve the equation $x^3 + x^2 = a$, a difficult problem to attack directly,

³VAT stands for Vorderasiatisches (Near East) Museum, in Berlin, part of the Pergamon Museum.

being in principle just as difficult as solving the general cubic equation. However, that was probably not the purpose of the tablet, which remains a mystery.

Neugebauer (1935, p. 99; 1952, p. 43) reports that the Mesopotamian mathematicians moved beyond algebra proper and investigated the laws of exponents, compiling tables of successive powers of numbers and determining the power to which one number must be raised in order to yield another. Such problems occur in a commercial context, involving compound interest. For example, the tablet AO 6484 gives the sum of the powers of 2 from 0 to 9 as the last term plus one less than the last term, as well as the sum of the squares of the first segment of integers as the sum of the same integers multiplied by the sum of the number $\frac{1}{3}$ and $\frac{2}{3}$ of the last integer in the segment. This recipe is equivalent to the modern formula for the sum of the squares of the first n integers. That is,

$$\sum_{k=1}^n k^2 = \frac{2n+1}{3} \left(\sum_{k=1}^n k \right).$$

PROBLEMS AND QUESTIONS

Mathematical Problems

- 4.1. Find the product $37; 11$, 7×6 , 13 ; 41 through standard multiplication.
- 4.2. Find two numbers whose sum is 15 and whose product is 40.25 by following the standard Mesopotamian technique of forming the average and semidifference.
- 4.3. Find two numbers whose difference is 4 and whose product is $\frac{76}{9}$ using the standard Mesopotamian technique.

Historical Questions

- 4.4. In what sense did the Mesopotamian authors “do algebra”? Did they have the concept of an equation or a classification of types of equations?
- 4.5. Give an example of a mathematical technique developed in Mesopotamia and extended to solve problems more general than the model on which it is based.
- 4.6. How did the Mesopotamian mathematicians deal with irrational square roots?

Questions for Reflection

- 4.7. For what purpose is it important to be able to find two numbers given their sum and product? Is there any practical application of this technique in everyday life?
- 4.8. For what purpose might a person need a table giving the sum of the cube and square of various numbers? Obviously, this table had *some* purpose, but was it perhaps simply an exercise in arithmetic for a pupil learning how to calculate? The expression can be interpreted geometrically, but does this geometric interpretation suggest any application?

- 4.9. The power of modern mathematical methods is so fascinating that there is a strong temptation to apply them to ancient texts. With our algebraic notation, we can reduce every cubic equation $ax^3 + bx^2 + cx + d = 0$ to an equation of the form $y^3 + y^2 = A$. Give at least two reasons why it is not plausible that the table of such values found on the tablet VAT 8402 had this purpose. [*Hint*: The transformation that brings about this reduction is the fractional-linear substitution $x \longrightarrow \frac{3b(3ac-b)y + (9abc - 2b^3 - 27a^2d)}{9a(b^2 - 3ac)y}$. Here $A = \frac{(9abc - 2b^3 - 27a^2d)^2}{(3(b^2 - 3ac))^3}$. (If you feel like verifying this fact, a computer algebra program such as *Mathematica* will help.)]

Geometry in Mesopotamia

Mesopotamian geometry was mostly concerned with the measurement of length, area, and volume. Still, many of the problems that are posed in geometric garb have no apparent practical application but are very good exercises in computation. For example, the Old Babylonian tablet BM 13901 contains the following problem: *Given two squares such that the side of one is two-thirds that of the other plus 5 GAR and whose total area is 25,25 square GAR, what are the sides of the squares?* Where in real life would one encounter such a problem? The tablet itself gives no practical context, and we conclude that this apparently geometric problem is really a computational problem. Earlier historians may have carried this idea too far. Neugebauer (1952, p. 41) stated, “It is easy to show that geometrical concepts play a very secondary part in Babylonian algebra, however extensively a geometrical terminology may be used.” Both Neugebauer and van der Waerden (1963, p. 72) point out that the cuneiform tablets contain operations that are geometrically absurd, such as adding a length to an area or multiplying two areas. These two giant figures in the history of mathematics a half century ago may have been too eager to press modern notions down on documents from the past. The very use of the word *algebra* tends to be misleading, since it suggests manipulation of symbols rather than numbers and the writing of equations, both of which are absent from the cuneiform tablets. Høyrup, one of the leading experts in this area, denies that any such dimensional inconsistency occurs, stating (2010, p. 5) that “No Babylonian text ever adds a number and either a length or an area.” His research (see Robson, 2009, p. 7), confirms that the numbers found in the geometric tablets are what we call concrete numbers, having a physical dimension like length or area.

5.1. THE PYTHAGOREAN THEOREM

There is conclusive evidence that the Mesopotamians knew the Pythagorean theorem at least 1000 years before Pythagoras (who, as we shall see, may have had nothing at all to do with it). They were thus already on the road to finding more abstract properties of geometric figures than mere size. This theorem was known at an early date in India and China, so that one cannot say certainly where the earliest discovery was and whether the appearance of this theorem in different localities was the result of independent discovery or transmission. But as far as present knowledge goes, the earliest examples of the use of the

“Pythagorean” principle that the square of the hypotenuse of a right triangle equals the sum of the squares of the other two legs occur in the cuneiform tablets. The Old Babylonian text known as BM 85196 contains a problem that has appeared in algebra books for centuries, the “leaning-ladder” problem, in which a ladder 30 units long is leaning against a wall, its top being 6 units below where it would be if pressed flush against the wall. The student is supposed to find how far away from the wall the bottom of the ladder is. In this problem we are dealing with a right triangle of hypotenuse 30 with one leg equal to $30 - 6 = 24$. Obviously, this is the famous 3–4–5 right triangle with all sides multiplied by 6. Obviously also, the interest in this theorem was more numerical than geometric. How often, after all, are we called upon to solve problems of this type in everyday life?

How might the Pythagorean theorem have been discovered? The following hypothesis was presented by Allman (1889, pp. 35–37), who cited a work (1870) by Carl Anton Bretschneider (1808–1878). Allman thought this dissection was due to the Egyptians, since, he said, it was done in their style.

Suppose that you find it necessary to construct a square twice as large as a given square. How would you go about doing so? (This is a problem the Platonic Socrates poses in the dialogue *Meno*.) You might double the side of the square, but you would soon realize that doing so actually quadruples the size of the square. If you drew out the quadrupled square and contemplated it for a while, you might be led to join the midpoints of its sides in order, that is, to draw the diagonals of the four copies of the original square. Since these diagonals cut the four squares in half, they will enclose a square twice as big as the original one (Fig. 5.1a). It is quite likely that someone, either for practical purposes or just for fun, discovered this way of doubling a square. If so, someone playing with the figure might have considered the result of joining in order the points at a given distance from the corners of a square instead of joining the midpoints of the sides. Doing so creates a square in the center of the larger square surrounded by four copies of a right triangle whose hypotenuse equals the side of the center square (Fig. 5.1b); it also creates the two squares on the legs of that right triangle and two rectangles that together are equal in area to four copies of the triangle. (In Fig. 5.1b, one of these rectangles is divided into two equal parts by its diagonal, which is the hypotenuse of the right triangle.) Hence the larger square consists of four copies of the right triangle plus the center square. It also consists of four copies of the right triangle plus the squares on the two legs of the right triangle. The inevitable conclusion is that *the square on the hypotenuse of any right triangle equals the sum of the squares on the legs*. This is the Pythagorean theorem, and it is used in many places in the cuneiform texts.

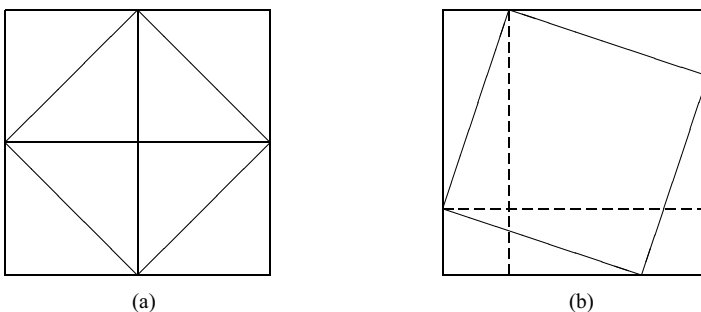


Figure 5.1. (a) Doubling a square; (b) the Pythagorean theorem.

Given that they knew the Pythagorean principle and also the polarization identity that makes it possible to express the product of two numbers as a difference of squares, it seems remarkable that the Mesopotamian mathematicians did not combine the two. If the hypotenuse of a right triangle is the average of two lengths and one of the legs is the semidifference, then the square on the other leg is the difference of the squares of the average and semidifference and, hence, is equal to the rectangle on the original two lengths. In this way, one can turn any rectangle into a square. The ingredients of a tasty mathematical dish were all there, but it does not appear that the Mesopotamians combined them and made them into a meal. It was left to the early Greeks to do that.

5.2. PLANE FIGURES

Some cuneiform tablets give the area of a circle in a way that we would interpret as implying $\pi = 3$. That statement, however, may mislead, since the procedure used for finding the area was not to multiply the square on the radius by a number, as we do, but to divide the square on the circumference by a number. That divisor needs to be 4π in our terms, and it appears to be 12 on at least one tablet. Hence the misleading shorthand that $\pi = 3$. On the other hand (Neugebauer, 1952, p. 46), the ratio of the circumference to the diameter, which we are going to call *one-dimensional* π ,¹ was given with more precision. On a tablet excavated at Susa in 1936, it was stated that the perimeter of a regular hexagon, which is three times its diameter, is 0 ; 57 , 36 times the circumference of the circumscribed circle. That makes the circumference of a circle of unit diameter equal to

$$\frac{3}{0 ; 57 , 36} = 3 ; 7 , 30 = 3 \frac{1}{8}.$$

That the Mesopotamian mathematicians saw a relation between the area and the circumference of a circle is shown by two Old Babylonian tablets from the Yale Babylonian Collection (YBC 7302 and YBC 11120, see Robson, 2001, p. 180). The first contains a circle with the numbers 3 and 9 on the outside and 45 on the inside. These numbers fit perfectly the formula $A = C^2/(4\pi)$, given that the scribe was using $\pi = 3$. Assuming that the 3 represents the circumference, 9 its square, and 45 the quotient, we find $9/(4 \cdot 3) = 3/4 = 0;45$. Confirmation of this hypothesis comes from the other tablet, which contains 1;30 outside and 11;15 inside, since $(1;30^2)/(4 \cdot 3) = (2;15)/12 = 135/12 = 11.25 = 11;15$.

5.2.1. Mesopotamian Astronomy

The strongest area of Mesopotamian science that has been preserved is astronomy, and it is here that geometry becomes most useful. The measurement of angles—arcs of circles—is essential to observation of the sun, moon, stars, and planets, since to the human eye they all appear to be attached to a large sphere rotating overhead. The division of a circle into

¹We use this term to specify the ratio of the circumference to the diameter of a circle. The geometric fact that this ratio is the same for all circles is taken for granted. What we shall call *two-dimensional* π is the ratio of the area enclosed by a circle to the interior of the square on its radius. Again, it is taken for granted that this ratio is the same for all circles. The reader can formulate the definition of three-dimensional π . That these ratios are all the same *real number* is not obvious, but seems to have been known even by ancient peoples.

360 degrees is one convention that came from Mesopotamia, was embraced by the Greeks, and became an essential part of applied geometry down to the present day. The reason for the number 360 is the base-60 computational system used in Mesopotamia. The astronomers divided all circles into 360 or 720 equal parts and divided the radius into 60 equal parts. In that way, a unit of length along the radius was approximately equal to a unit of length on the circle.

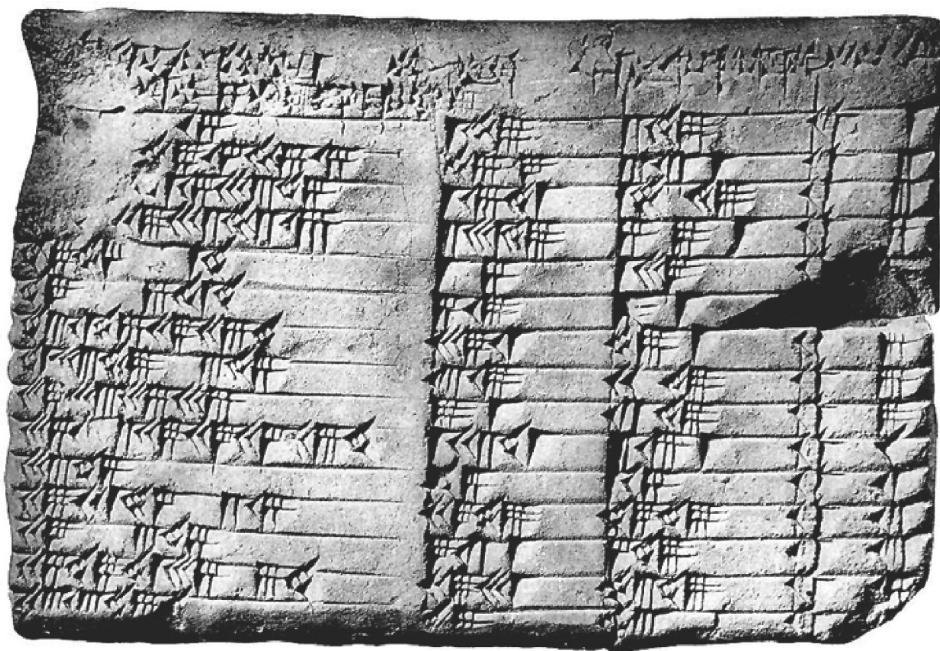
5.3. VOLUMES

The cuneiform tablets contain computations of the volumes of some simple solid figures. For example, the volume of a frustum of a square pyramid is computed in an Old Babylonian tablet (BM 85194). The Mesopotamian scribe seems to have generalized incorrectly from the case of a trapezoid and reasoned that the volume is the height times the average area of the upper and lower faces. This rule overestimates the volume. There is, however, some disagreement as to the correct translation of the tablet in question. Neugebauer (1935, Vol. 1, p. 187) claimed that the computation was based on an algebraic formula that is geometrically correct. The square bases are given as having sides 10 and 7, respectively, and the height is given as 18. The incorrect rule just mentioned would give a volume of 1341, which is 22,21 in sexagesimal notation; but the actual text reads 22,30. The discrepancy could be a simple misprint, with three ten-symbols carelessly written for two ten-symbols and a one-symbol. The computation used is not entirely clear. The scribe first took the average base side $(10 + 7)/2$ and squared it to get 1, 12; 15 in sexagesimal notation (72.25). At this point there is apparently some obscurity in the tablet itself. Neugebauer interpreted the next number as 0; 45, which he assumed was calculated as one-third of the square of $(10 - 7)/2$. The sum of these two numbers is 1, 13, which, multiplied by 18, yields 21, 54 (that is, 1314), which is the correct result. But it is difficult to see how this number could have been recorded incorrectly as 22, 30. If the number that Neugebauer interprets as 0; 45 is actually 2; 15 (which is a stretch—three ten-symbols would have to become two one-symbols), it would be exactly the square of $(10 - 7)/2$, and it would yield the same incorrect formula as the assumption that the average of the areas of the two bases was being taken. In any case, the same procedure is used to compute the volume of the frustum of a cone (Neugebauer, 1935, p. 176), and in that case it definitely is the incorrect rule stated here, taking the average of the two bases and multiplying by the height.

5.4. PLIMPTON 322

The diagonals and sides of rectangles are the subject of a cuneiform tablet from the period 1900–1600 BCE, number 322 of the Plimpton collection at Columbia University. The numbers on this tablet have intrigued many mathematically oriented people, leading to a wide variety of speculation as to the original purpose of the tablet.

As you can see from the photograph, there are a few chips missing, so that some of the cuneiform numbers in the tablet will need to be restored by plausible conjecture. Notice also that the column at the right-hand edge contains the cuneiform numbers in the sequence 1, 2, 3, 4, . . . , . . . , 7, 8, 9, 10, 11, 12, 13, . . . , Obviously, this column merely numbers the rows. The column second from the right consists of identical symbols that we shall ignore entirely. Pretending that this column is not present, if we transcribe only what we can see



Plimpton 322. © Rare Book and Manuscript Library, Columbia University.

into our version of sexagesimal notation, denoting the chipped-off places with brackets ([. . .]), we get the four-column table shown below.

Before analyzing the mathematics of this table, we make one preliminary observation: Row 13 is anomalous, in that the third entry is smaller than the second entry. For the time being, we shall ignore this row and see if we can figure out how to correct it. Row 15 (the bottom row) is damaged, and we shall temporarily exclude it from consideration. Since the long numbers in the first column must be the result of computation—it is unlikely that measurements could be carried out with such precision—we make the reasonable conjecture that the shorter numbers in the second and third columns are data. As mentioned in the preceding chapter, the Mesopotamian mathematicians routinely associated with any pair of numbers (a, b) two other numbers: their average $(a + b)/2$ and their semidifference $(b - a)/2$. Let us compute these numbers for all the rows except rows 13 and 15 to see how they would have appeared to a mathematician of the time. We get the following 13 pairs of numbers, which we write in decimal notation: (144, 25), (7444, 4077), (5625, 1024), (15625, 2916), (81, 16), (400, 81), (2916, 625), (1024, 225), (655, 114), (6561, 1600), (60, 15), (2304, 625), (2500, 729).

You will probably recognize a large number of perfect squares in the table. Indeed, *all* of these numbers, except for those corresponding to rows 2, 9, and 11, are perfect squares: 10 pairs of perfect squares out of thirteen! That is too unusual to be a mere coincidence. A closer examination reveals that they are squares of numbers whose only prime factors are 2, 3, and 5. Now these are precisely the prime factors of the number 60, which the Mesopotamian mathematicians used as a base. That means that the reciprocals of these numbers will have terminating sexagesimal expansions. We should therefore keep in mind that the reciprocals of these numbers may play a role in the construction of the table.

Notice also that these ten pairs are all *relatively prime* pairs. Let us now denote the square root of the average by p and denote the square root of the semidifference by q . Column 2 will then be $p^2 - q^2$, and column 3 will be $p^2 + q^2$. Having identified the pairs (p, q) as important clues, we now ask *which* pairs of integers occur here and how they are arranged. The values of q , being smaller, are easily handled. The smallest q that occurs is 5 and the largest is 54, which also is the largest number less than 60 whose only prime factors are 2, 3, and 5. Thus, we could try constructing such a table for all values of q less than 60 having only those prime factors. But what about the values of p ? Again, ignoring the rows for which we do not have a pair (p, q) , we observe that the rows occur in decreasing order of p/q , starting from $12/5 = 2.4$ and decreasing to $50/27 = 1.85185185\dots$. Let us then impose the following conditions on the numbers p and q :

	Width	Diagonal	
[...] 15	1,59	2,49	1
[...] 58,14,50,6,15	56,7	3,12,1	2
[...] 41,15,33,45	1,16,41	1,50,49	3
[...] 29,32,52,16	3,31,49	5,9,1	4
48,54,1,40	1,5	1,37	5
47,6,41,40	5,19	8,1	6
43,11,56,28,26,40	38,11	59,1	7
41,33,45,14,3,45	13,19	20,49	8
38,33,36,36	9,1	12,49	9
35,10,2,28,27,24,26	1,22,41	2,16,1	10
33,45	45	1,15	11
29,21,54,2,15	27,59	48,49	12
27,[...],3,45	7,12,1	4,49	13
25,48,51,35,6,40	29,31	53,49	14
23,13,46,40	[...]	[...]	[...]

1. The integers p and q are relatively prime.
2. The only prime factors of p and q are 2, 3, and 5.
3. $q < 60$.
4. $1.8 \leq p/q \leq 2.4$

Now, following an idea of Price (1964), we ask which possible (p, q) satisfy these four conditions. We find that every possible pair occurs with only five exceptions: (2, 1), (9, 5), (15, 8), (25, 12), and (64, 27). There are precisely five rows in the table—rows 2, 9, 11, 13, and 15—for which we did not find a pair of perfect squares. Convincing proof that we are on the right track appears when we arrange these pairs in decreasing order of the ratio p/q . We find that (2, 1) belongs in row 11, (9, 5) in row 15, (15, 8) in row 13, (25, 12) in row 9, and (64, 27) in row 2, precisely the rows for which we did not previously have a pair p, q . The evidence is overwhelming that these rows were intended to be constructed using these pairs (p, q) . When we replace the entries that we can read by the corresponding numbers $p^2 - q^2$ in column 2 and $p^2 + q^2$ in column 3, we find the following:

In row 2, the entry 3,12,1 has to be replaced by 1,20,25, that is, 11521 becomes 4825. The other entry in this row, 56,7, is correct.

In row 9, the entry 9,1 needs to be replaced by 8,1, so here the writer simply inserted an extra unit character.

In row 11, the entries 45 and 75 must be replaced by 3 and 5; that is, both are divided by 15. It has been remarked that if these numbers were interpreted as $45 \cdot 60$ and $75 \cdot 60$, then in fact, one would get $p = 60$, $q = 30$, so that this row was not actually “out of step” with the others. But of course when that interpretation is made, p and q are no longer relatively prime, in contrast to all the other rows.

In row 13 the entry 7,12,1 must be replaced by 2,41; that is, 25921 becomes 161. In other words, the table entry is the square of what it should be.

The illegible entries in row 15 now become 56 and 106. The first of these is consistent with what can be read on the tablet. The second tablet entry appears to be 53, half of what it should be.

The final task in determining the mathematical meaning of the tablet is to explain the numbers in the first column and interpolate the missing pieces of that column. Notice that the second and third columns in the table are labeled “width” and “diagonal.” Those labels tell us that we are dealing with dimensions of a rectangle here and that we should be looking for its length. By the Pythagorean theorem, that length is $\sqrt{(p^2 + q^2)^2 - (p^2 - q^2)^2} = \sqrt{4p^2q^2} = 2pq$. Even with this auxiliary number, however, it requires some ingenuity to find a formula involving p and q that fits the entries in the first column that can be read. If the numbers in the first column are interpreted as the sexagesimal representations of numbers between 0 and 1, those in rows 5 through 14—the rows that can be read—all fit the formula²

$$\left(\frac{p/q - q/p}{2} \right)^2.$$

Assuming this interpretation, since it works for the 10 entries we can read,³ we can fill in the missing digits in the first four and last rows. This involves adding one or two digits to the beginning of the first four rows, and it appears that there is just the right amount of room in the chipped-off place to allow this to happen.⁴ The digits that occur in the bottom

²In some discussions of Plimpton 322 the claim is made that a sexagesimal 1 should be placed before each of the numbers in the first column. Although the tablet is clearly broken off on the left, it does not appear from pictures of the tablet—the author has never seen it “live”—that there were any such digits there before. Neugebauer (1952, p. 37) claims that parts of the initial 1 remain from line 4 on “as is clearly seen from the photograph” and that the initial 1 in line 14 is completely preserved. When that assumption is made, however, the only change in the interpretation is a trivial one: The negative sign in the formula must be changed to a positive sign, and what Friberg interpreted as a column of squares of tangents becomes a column of squares of secants, since $\tan^2 \theta + 1 = \sec^2 \theta$.

³Comparing the original in the photograph with the computed values in the first column of the table, it appears that the original tablet has the single digit 59 in the middle of Column 1 of row 8 instead of the computed digits 45, 14 which we entered in the table. This seeming discrepancy—if it is a discrepancy—is easily explained by assuming that the scribe simply merged the two sexagesimal digits, which have a total of five 10-symbols and nine 1-symbols between them.

⁴The digits to be inserted are as follows. Row 1: 59, 0 (but the zero would have been a blank space in the original). Row 2: 56, 56. Row 3: 55, 7. Row 4: 53, 10.

row are 23,13,46,40, and they are consistent with the parts that can be read from the tablet itself.

5.4.1. The Purpose of Plimpton 322: Some Conjectures

The *structure* of the tablet is no longer a mystery, except for the tiny mystery of the misprint in row 2, column 3. Its *purpose*, however, is not clear. What information was the table intended to convey? Was it intended to be used as people once used tables of products, square roots, and logarithms—that is, to look up a number or pair of numbers? If so, which columns contained the input and which the output? One geometric problem that can be solved by use of this tablet is that of multiplying a square by a given number; that is, given a square of side a , it is possible to find the side b of a square whose ratio to the first square is given in the first column. To do so, take a rope whose length equals the side a and divide it into the number of equal parts given in the second column, then take a second rope with the same unit of length and total length equal to the number of units in the third column and use these two lengths to form a leg and the hypotenuse of a right triangle. The other leg will then be the side of a square having the given ratio to the given square. The problem of shrinking or enlarging squares was considered in other cultures, but such an interpretation of Plimpton 322 has only the merit that there is no way of proving the tablet *wasn't* used in this way. There is no proof that the tablet was ever put to this use.

Friberg (1981) suggested that the purpose of the tablet was trigonometrical—that is, that it was a table of squares of tangents. Columns 2 and 3 give one leg and the hypotenuse of 15 triangles with angles intermediate between those of the standard 45–45–90 and 30–60–90 triangles. What is very intriguing is that, if this was its purpose, the table covers the case of all possible triangles whose shapes are between these two and whose legs have lengths that are multiples of a standard unit by numbers having only 2, 3, and 5 as factors. Of all right triangles, the 45–45–90 and the 30–60–90 are the two that play the most important role in all kinds of geometric applications; plastic models of them were once used as templates in mechanical drawing, and such models are still sold. It is easy to imagine that a larger selection of triangle shapes might have been useful in the past, before modern drafting instruments and computer-aided design. Using this table, one could build 15 model triangles with angles varying in increments of approximately 1° . One can imagine such models being built and the engineer of 4000 years ago reaching for a “number 7 triangle” when a slope of $574/675 = .8504$ was needed. However, this scenario still lacks plausibility. Even if we assume that the engineer kept the tablet around as a reference when it was necessary to know the slope, the tablet stores the *square* of the slope in column 1. It is difficult to imagine any engineering application for that number.

We now explore the computational and pedagogical possibilities inherent in this tablet. The left-hand column contains numbers that are perfect squares and remain perfect squares when 1 is added to them. If the purpose of the table was to generate numbers with this property, the use of the table would be as follows: Square the entry in column 3, square the entry in column 2, and then divide each by the difference of these squares. The results of these two divisions would be two squares differing by 1. The numbers p and q that generate the two columns can be arbitrary, but in order to get a sexagesimally terminating entry in the first column, the difference $(p^2 + q^2)^2 - (p^2 - q^2)^2 = 4p^2q^2$ should have only 2, 3, and 5 as prime factors, and hence p and q also should have only these factors. Against this interpretation there lies the objection that p and q are concealed from the casual reader of the tablet.

In a paper that was never published (see Buck, 1980, p. 344), D. L. Voils⁵ pointed out that tablets amounting to “teacher’s manuals” have been found in which the following problem is set: *Find a number that yields a given number when its reciprocal is subtracted.* In modern terms, this problem requires solving the equation

$$x - \frac{1}{x} = d,$$

where d is the given number. Obviously, if you were a teacher setting such a problem for a student, you would want the solution x to be such that both x and $1/x$ have terminating sexagesimal digits. So, if the solution is to be $x = p/q$, we already see why we need both p and q to be products of 2, 3, and 5. This problem amounts to the quadratic equation $x^2 - dx - 1 = 0$, and its unique positive solution is $x = d/2 + \sqrt{1 + (d/2)^2}$. Column 1 of the tablet, which contains $(d/2)^2$, then appears as part of the solution process. It is necessary to take its square root and also the square root of $\sqrt{1 + d^2/4}$ (which is the same number with a 1 prefixed to it) in order to find the solution $x = p/q$. This explanation seems to fit very well with the tablet. One could assume that the first column gives values of d that a teacher could use to set such a problem with the assurance that the pupil would get terminating sexagesimal expansions for both x and $1/x$. This hypothesis also fits very well with the fact already noted that many of the cuneiform tablets are pedagogical in nature. On the other hand, it does not fully explain why the tablet gives the numbers $p^2 - q^2$ and $p^2 + q^2$, rather than simply p and q , in subsequent columns. Doing our best for this theory, we note that columns 2 and 3 contain, respectively, the numerators of $x - 1/x$ and $x + 1/x$ and that their common denominator is the square root of the difference of the squares of these two numerators. Against that explanation is the fact that the Mesopotamians did not work with common fractions. The concepts of numerator and denominator to them would have been the concepts of dividend and divisor, and the final sexagesimal quotient would not display these numbers. Still, a terminating sexagesimal expansion is a common fraction, and these special numbers (the only numbers ever considered in Mesopotamia) amount to the use of common fractions, confined to those whose denominators are products of powers of 2, 3, and 5. The recipe for getting from columns 2 and 3 to column 1 would be first to square each of these columns, then find the reciprocal of the difference of the squares as a sexagesimal expansion, and, finally, multiply the last result by the square in column 2.

PROBLEMS AND QUESTIONS

Mathematical Problems

- 5.1.** Explain the author’s solution of the “leaning-ladder” problem from the cuneiform tablet BM 85196. Here the numbers in square brackets were worn off the tablet and have been reconstructed.

⁵In early 2011, Prof. Douglas Rogers of the University of Hawaii made a diligent search and located Voils, now retired in Florida. Voils reported that he had indeed written such a paper, but declined to revise it for publication, since his interest had shifted to computer science.

A beam of length 0;30 GAR [about 3 meters] is leaning against a wall. Its upper end is 0;6 GAR lower than it would be if it were perfectly upright. How far is its lower end from the wall?

Do the following: Square 0;30, obtaining 0;15. Subtracting 0;6 from 0;30 leaves 0;24. Square 0;24, obtaining 0;9,36. Subtract 0;9,36 from [0;15], leaving 0;5,24. What is the square root of 0;5,24? The lower end of the beam is [0;18] from the wall.

When the lower end is 0;18 from the wall, how far has the top slid down? Square 0;18, obtaining 0;5,24. . . .

- 5.2. Show that the average of the areas of the two bases of a frustum of a square pyramid is the sum of the squares of the average and semidifference of the sides of the bases.
- 5.3. Solve the problem from BM 13901, finding two squares, one of which has a side five units longer than the other and whose total area is 25 , 25 square units.

Historical Questions

- 5.4. From which time period and dynasty do the Old Babylonian tablets come?
- 5.5. What standard geometric figures are studied in the Old Babylonian tablets discussed in this chapter?
- 5.6. Why is it misleading to talk about the “Babylonian value of π ”?

Questions for Reflection

- 5.7. In what everyday applications might some of the geometric problems discussed above (such as finding the volume of a frustum of a pyramid) be useful?
- 5.8. Could the relation noted above between the areas of the two bases of a frustum of a square pyramid and the squares of the average and semidifference of their sides have led the Mesopotamian mathematicians astray in their computation of the volume of the frustum?
- 5.9. Given that it is difficult to think of applications of the many geometric problems studied in the tablets, what could have been the motive for writing them?

Egyptian Numerals and Arithmetic

The earliest systematic treatises on mathematics come from the Egyptian civilization, which was already 2000 years old before the mathematical treatises that survive today were written. After several thousand years during which the area now called Egypt was the home of isolated agricultural communities, a process of consolidation began, and by 3100 BCE there were two major kingdoms, Upper Egypt in the south and Lower Egypt in the north. Egypt became politically unified about this time when a ruler of Upper Egypt, variously said to be named Menes, Narmer, or “Scorpion,” conquered Lower Egypt. In the four centuries following this conquest, a number of technological advances were made in Egypt, making it possible to undertake large-scale engineering projects. Such projects required a certain amount of arithmetic and geometry. Shortly after the beginning of the Old Kingdom (2685 BCE) the famous Step Pyramid of Djoser was built, the first structure made entirely of hewn stone. The Old Kingdom, which lasted just over five centuries, was a time of active building of temples and tombs. The collapse of central authority at the end of this period led to a century and a half during which the real power was held by provincial governors. The central authority recovered when the governors of Thebes extended their power northward and over several generations brought about the Middle Kingdom (2040–1785 BCE).

When the central authority weakened again at the end of this period, foreign invaders known as the Hyksos conquered most of Egypt from the north. The Hyksos rule lasted for about a century, until some of their puppet governors became strong enough to usurp their authority; the Hyksos were driven out in 1570 BCE, which marked the beginning of the New Kingdom. It was during the Hyksos period that the earliest mathematical treatises still extant were written. We therefore begin with a discussion of mathematics as practiced in the Middle Kingdom.

6.1. SOURCES

Mathematics has been practiced in Egypt continuously starting at least 4000 years ago. Aristotle believed that the study of mathematics first arose among the Egyptians. In his *Metaphysics* (Bekker¹ 981b), he wrote

¹Analogous to the Stephanus indexing of the works of Plato, which will be mentioned in Section 8.2 of Chapter 8, the works of Aristotle were issued by the Prussian Academy of Sciences in the nineteenth century, edited by August Immanuel Bekker (1785–1871).

Thus it was that the mathematical sciences first arose in Egypt. For it was there that the priestly caste was granted the necessary leisure.

In the late fourth century BCE, it merged with the mathematics of the Greeks, who had learned the basics of geometry from the Egyptians. Indeed, the intellectual center of the Western world, the city of Alexandria founded by Alexander the Great, was in Egypt. Centuries later, Egypt formed part of the Muslim culture centered in Baghdad, which also produced some brilliant mathematics, and mathematical creativity continues in Egypt at the present day. The Egyptian mathematics we are going to discuss, however, had a beginning and an end. It began with hieroglyphic inscriptions containing numbers and dating to the third millennium BCE and ended at the time of Euclid, in the third century BCE. After that time, the city of Alexandria in the Nile delta was the main school of mathematics in the Hellenistic world, and many of the most prominent mathematicians who wrote in Greek studied there. We shall confine our attention, however, to what was for many centuries the standard set of mathematical techniques used by the professionals who administered the Egyptian state during the Middle Kingdom.

6.1.1. Mathematics in Hieroglyphics and Hieratic

The great architectural monuments of ancient Egypt are covered with hieroglyphic characters, some of which contain numbers. In fact, the ceremonial mace of the founder of the first dynasty contains records that mention oxen, goats, and prisoners and contain hieroglyphic symbols for the numbers 10,000, 100,000, and 1,000,000. These hieroglyphic symbols, although suitable for ceremonial recording of numbers, were not well adapted for writing on papyrus or leather. The language of the earliest written documents that have been preserved to the present time is a cursive known as *hieratic*.

The most detailed information about Egyptian mathematics comes from a single document written in the hieratic script on papyrus around 1650 BCE and preserved in the dry Egyptian climate. This document is known as the Rhind papyrus, after the British lawyer Alexander Rhind (1833–1863), who went to Egypt for his health and became an Egyptologist. Rhind purchased the papyrus in Luxor, Egypt, in 1857. Parts of the original document have been lost, but a section consisting of 14 sheets glued end to end to form a continuous roll $3\frac{1}{2}$ feet wide and 17 feet long remains. Part of it is on public display in the British Museum, where it has been since 1865. Some missing pieces of this document were discovered in 1922 in the Egyptian collection of the New York Historical Society; these are now housed at the Brooklyn Museum of Art. A slightly earlier mathematical papyrus, now in the Moscow Museum of Fine Arts, consists of sheets about one-fourth the size of the Rhind papyrus. This papyrus was purchased by V. S. Golenishchev (1856–1947) in 1893 and donated to the museum in 1912. A third document, a leather roll purchased along with the Rhind papyrus, was not unrolled for 60 years after it reached the British Museum because the curators feared it would disintegrate if unrolled. It was some time before suitable techniques were invented for softening the leather, and the document was finally unrolled in 1927. The contents turned out to be a collection of 26 sums of unit fractions, from which historians were able to gain insight into Egyptian methods of calculation. A fourth set of documents, known as the Reisner papyri after the American archaeologist George Andrew Reisner (1867–1942), who purchased them in 1904, consists of four rolls of records from dockyard workshops, apparently from the reign of Senusret I (1971–1926 BCE). They are now in the Boston Museum of Fine Arts. Another document, the Akhmim Wooden Tablet,

is housed in the Egyptian Museum in Cairo. The Akhmim Wooden Tablet contains several ways of expressing reciprocals of integers based on dividing unity ($64/64$) by these integers. According to Milo Gardner (<http://mathworld.wolfram.com/AkhmimWoodenTablet.html>), the significance of the number 64 is that it is the number of *ro* in a *hekat* of grain. It also relates to the so-called *Horus-eye fractions*, as we shall discuss below. This origin for the numbers makes sense and gives a solid practical origin for Egyptian arithmetic. These documents show the practical application of Egyptian mathematics in construction and commerce. We shall mostly discuss the Rhind papyrus in this chapter and the next, only occasionally mentioning items from the others.

6.2. THE RHIND PAPYRUS

What do these documents tell us about the practice of mathematics in ancient Egypt? The author of the Rhind papyrus begins his work by describing it as a “correct method of reckoning, for grasping the meaning of things, and knowing everything that is, obscurities. . . and all secrets.”² The author seems to value mathematics because of its explanatory power, but that explanatory power was essentially practical, not at all mystical.

We are fortunate to be able to date the Rhind papyrus with such precision. The author, a scribe named Ahmose (or Ahmes), gives us his name and tells us that he is writing in the fourth month of the flood season of the thirty-third year of the reign of Pharaoh A-user-re (Apepi I). From this information, Egyptologists arrived at a date of around 1650 BCE for this papyrus, which is approximately the latest date of the Old Babylonian cuneiform tablets discussed in the preceding chapters. Ahmose tells us, however, that he is merely copying work written down in the reign of Pharaoh Ny-maat-re, also known as Amenemhet III (1842–1797 BCE), the sixth pharaoh of the Twelfth Dynasty. From that information it follows that the mathematical knowledge contained in the papyrus is nearly 4000 years old.

The introductory paragraph of the Rhind papyrus is followed by certain tables that resemble multiplication tables (more on this subject below), along with 87 problems involving various mathematical processes. Attempts have been made to discern a pattern in the arrangement of these problems. The suggestion that seems most plausible intuitively is that the problems are grouped according to their application rather than their method of solution. The first six problems, for example, involve dividing loaves of bread among 10 people. Problems 7–23 are purely technical and show how to add fractional parts and, given a certain number of fractional parts, how to find complementary fractional parts to obtain a whole. Problems 24–38 are concerned with finding a quantity of which certain fractional parts will yield a given number. Area, volume, and general measurement problems are numbered from 40 to 60, and the remaining problems are concerned with various commercial applications to the distribution of goods.

6.3. EGYPTIAN ARITHMETIC

Let us begin our discussion of Egyptian mathematics by describing the way numbers were written. In hieroglyphics, numbers are represented as vertical strokes (|) for each individual

²This is the translation given by Robins and Shute (1987, p. 11). Chace et al. (1927, p. 49) give the translation as “the entrance into the knowledge of all existing things and all secrets.”

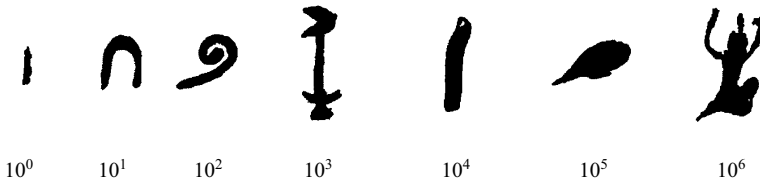


Figure 6.1. Powers of 10 from 10^0 to 10^6 in hieroglyphics.

digit, up to 9; then 10 is written as \cap , 20 as $\cap\cap$, and so on. To represent 100, the Egyptians used a symbol resembling a coil of rope. Such a system requires new symbols to be invented for higher and higher groupings, as larger and larger numbers become necessary. As Fig. 6.1 shows, the Egyptians had hieroglyphic symbols for 1000 (a lotus blossom), 10,000 (a crooked thumb), 100,000 (a turbot fish), and 1,000,000 (said to be the god of the air). With this system of recording numbers, no symbol for zero was needed, nor was the order of digits of any importance, since, for example, $||| \cap \cap$ and $\cap \cap |||$ both mean 23. The disadvantage of the notation is that the symbol for each power of 10 must be written a number of times equal to the digit that we would put in its place. This system is very far from our place-value decimal system, in which there are symbols for the numbers 0 through 9, interpreted as multiples of powers of ten, with the power indicated by physical location in the number. Later on, in the hieratic script that replaced hieroglyphics, they had special symbols for 1 through 9, 10 through 90, 100 through 900, and so on, as shown in Fig. 6.2. This system was later used by the Greeks, with Greek letters replacing the hieratic symbols.

6.4. COMPUTATION

After the descriptive title, the Rhind papyrus exhibits the table of numbers shown in Fig. 6.3, which will be discussed below. In contrast to our modern arithmetic, which consists of the four operations of addition, subtraction, multiplication, and division performed on whole numbers and fractions, the fundamental operations in Egypt were addition, subtraction, and *doubling*, and these operations were performed on whole numbers and *parts*. We need to discuss both the operations and the objects on which they were carried out.

Let us consider first the absence of multiplication and division as we know them. First of all, there is something special about the number 2. It is a number that we can grasp easily without even having to count. It has always played a special role in ordinary conversation. For example, we don't normally say "one-twoth" for the result of dividing something in two parts. This linguistic peculiarity suggests that *doubling* is psychologically different from applying the general concept of multiplying in the special case when the multiplier is 2.

Next consider the absence of what we call fractions. The closest Egyptian equivalent to a fraction is something we shall call a *part*. For example, what we refer to nowadays as the fraction $\frac{1}{7}$ would be referred to as "the seventh part." This way of expressing fractions has a venerable history, even in English, and you will frequently encounter it in writing from earlier centuries. The phrase "seventh part" conveys the image of a thing divided into seven equal parts arranged in a row and the seventh (and last) one being chosen. For that reason, according to van der Waerden (1963), there can be only one seventh part, namely the last one; there would be no way of expressing what we call the fraction $\frac{3}{7}$, since there




























	1	10	100
1			
2			
3			
4			
5			
6			
7			
8			
9			

Figure 6.2. Hieratic symbols, arranged as a multiplication table.

couldn't be three seventh parts. An exception was the fraction that we call $\frac{2}{3}$, which occurs constantly in the Rhind papyrus. There was a special symbol meaning "the two parts" out of three. It is very easy to interpret parts in our own language. They are *unit fractions*, that is, fractions whose numerator is 1. But for historical purposes, it is better to retain the obsolete language of parts. Our familiarity with fractions in general makes it difficult to see what the fuss is about when the author asks what must be added to the two parts and the fifteenth part in order to make a whole (Problem 21 of the papyrus). If this problem is stated in modern notation, it merely asks for the value of $1 - (\frac{1}{15} + \frac{2}{3})$. We get the answer immediately, expressing it as $\frac{4}{15}$. Both this process and the answer would have been foreign to the Egyptian, whose solution is described below.

To understand the Egyptians, we shall try to imitate their way of writing down a problem. On the other hand, we would be at a great disadvantage if our desire for authenticity led us to try to solve the entire problem using their notation. The best compromise seems to

be to use our symbols for the whole numbers and express a *part* by the corresponding whole number with a bar over it. Thus, *the fifth part* will be written $\bar{5}$, *the thirteenth part* will be represented by $\bar{13}$, and so on. For “the two parts” ($\frac{2}{3}$) we shall use a double bar, that is, $\bar{\bar{3}}$.

6.4.1. Multiplication and Division

Since the only operation other than addition and subtraction of integers (which are performed automatically without comment) is doubling, the problem that we would describe as “multiplying 11 by 19” would have been written out as follows:

	19	1	*
	38	2	*
	76	4	
	152	8	*
Result	209	11	

Inspection of this process shows its justification. The rows are kept strictly in proportion by doubling each time. The final result can be stated by comparing the first and last rows: 19 is to 1 as 209 is to 11. The rows in the right-hand column that must be added in order to obtain 11 are marked with an asterisk, and the corresponding entries in the left-hand column are then added to obtain 209. In this way any two positive integers can easily be multiplied. The only problem that arises is to decide how many rows to write down and which rows to mark with an asterisk. But that problem is easily solved. You stop creating rows when the next entry in the right-hand column would be bigger than the number you are multiplying by (in this case 11). You then mark your last row with an asterisk, subtract the entry in its right-hand column (8) from 11 (getting a remainder of 3), then move up and mark the next row whose right-hand column contains an entry not larger than this remainder (in this case the second row), subtract the entry in its right-hand column (2), from the previous remainder to get a smaller remainder (in this case 1), and so forth.

We shall refer to this general process of doubling and adding as *calculating*. What we call division is carried out in the same way, by reversing the roles of the two columns. For example, what we would call the problem of dividing 873 by 97 amounts to calculating with 97 so as to obtain 873. We can write it out as follows:

*	97	1	
	194	2	
	388	4	
*	776	8	
	873	9	Result

The process, including the rules for creating the rows and deciding which ones to mark with an asterisk, is exactly the same as in the case of multiplication, except that now it is the left-hand column that is used rather than the right-hand column. We create rows until the next entry in the left-hand column would be larger than 873. We then mark the last row, subtract the entry in its left-hand column from 873 to obtain the remainder of 97, then look for the next row above whose left-hand entry contains a number not larger than 97, mark that row, and so on.

6.4.2. “Parts”

It has probably occurred to you that the second use of the two-column system may lead to complications. While in the first problem we can always express any positive integer as a sum of powers of 2, the second problem is a different matter. We were just lucky that we happened to find multiples of 97 that add up to 873. If we hadn’t found them, we would have had to deal with those *parts* that have already been discussed. For example, if the problem were “calculate with 12 so as to obtain 28,” it might have been handled as follows:

	12	1	
*	24	$\frac{2}{3}$	
	8	$\frac{2}{3}$	
*	4	$\frac{2}{3}$	
	28	$2\frac{2}{3}$	Result

What is happening in this computation is the following. We stop creating rows after 24 because the next entry in the left-hand column (48) would be bigger than 28. Subtracting 24 from 28, we find that we still need 4, yet no 4 is to be found. We therefore go back to the first row and multiply by $\frac{2}{3}$, getting the row containing 8 and $\frac{2}{3}$. Dividing by 2 again gets a 4 in the left-hand column. We then have the numbers we need to get 28, and the answer is expressed as $2\frac{2}{3}$. Quite often the first multiplication by a *part* involves the two-thirds part $\frac{2}{3}$. The scribes probably began with this part instead of one-half for the same reason that a carpenter uses a plane before sandpaper: The work goes faster if you take bigger “bites.”

The parts that are negative powers of 2 play a special role. When applied to a hekat of grain, they are referred to as the *Horus-eye* parts. According to Egyptian legend, the god Horus lost an eye in a fight with his uncle, and the eye was restored by the god Thoth. Each of these fractions was associated with a particular part of Horus’ eye. Since $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \frac{1}{64} = \frac{63}{64}$, the scribes apparently saw that unity could be restored (approximately), as Horus’ eye was restored, by using these parts. The fact that (in our terms) 63 occurs as a numerator shows that division by 3, 7, and 9 is facilitated by the use of the Horus-eye series. In particular, since $\frac{1}{7} = \frac{1}{7} \cdot \left(\frac{63}{64} + \frac{1}{64}\right) = \frac{9}{64} + \frac{1}{448} = \frac{8}{64} + \frac{1}{64} + \frac{1}{448}$, the seventh part could have been written as $\frac{8}{64} \frac{1}{448}$. In this way, the awkward seventh part gets replaced by the better-behaved Horus-eye fractions, plus a corrective term (in this case $\frac{1}{448}$, which might well be negligible in practice). Five such replacements are implied, though not given in detail, in the Akhmim Wooden Tablet. As another example, since $64 = 4 \cdot 13 + 8 + 4$, which, when both expressions are divided by 13×64 , becomes $\frac{1}{13} = \frac{1}{16} + \frac{1}{8 \times 13} + \frac{1}{16 \times 13}$, we find that $\frac{1}{13} = \frac{1}{16} \frac{1}{104} \frac{2}{208}$. This expansion makes it easy to see how to write the double of $\frac{1}{13}$ in terms of parts.

There are two more complications that arise in doing arithmetic the Egyptian way. The first complication is obvious. Since the procedure is based on doubling, but the double of a *part* may not be expressible as a part, how does one “calculate” with parts? It is easy to double, say, the twenty-sixth part: The double of the twenty-sixth part is the thirteenth part. If we try to double again, however, we are faced with the problem of doubling a part involving an odd number. The table at the beginning of the papyrus gives the answer: The double of the thirteenth part is the eighth part plus the fifty-second part plus the one hundred fourth part. In our terms this tabular entry expresses the fact that

$$\frac{2}{13} = \frac{1}{8} + \frac{1}{52} + \frac{1}{104}.$$

5	3 15	55	30 318 795
7	4 28	57	38 114
9	6 18	59	36 236 531
11	6 66	61	40 244 488 610
13	8 52 104	63	42 126
15	10 30	65	39 195
17	12 51 68	67	40 335 536
19	12 76 114	69	46 138
21	14 42	71	40 568 710
23	12 276	73	60 219 292 365
25	15 75	75	50 150
27	18 54	77	44 308
29	24 58 174 232	79	60 237 316 790
31	20 124 155	81	54 162
33	22 66	83	60 332 415 498
35	30 42	85	51 255
37	24 111 296	87	58 174
39	26 78	89	60 356 534 890
41	24 246 328	91	70 130
43	42 86 129 301	93	62 186
45	30 90	95	60 380 570
47	30 141 470	97	56 679 776
49	28 196	99	66 198
51	34 102	101	101 202 303 606

Figure 6.3. Doubles of unit fractions in the Rhind papyrus.

Gillings (1972, p. 49) lists five precepts apparently followed by the compiler of this table in order to make it maximally efficient for use. The most important of these are the following three. One would like each double (1) to have as few terms as possible, (2) with each term as small as possible (that is, the “denominators” as small as possible), and (3) with even “denominators” rather than odd ones. These principles have to be balanced against one another, and the table in Fig. 6.3 represents the resulting compromise. However, Gillings’ principles are purely negative ones, telling what *not* to do. The positive side of creating such a table is to find simple patterns in the numbers. One pattern that occurs frequently is illustrated by the double of $\bar{5}$, and amounts to the identity $2/p = 1/((p+1)/2) + 1/(p(p+1)/2)$. Another, illustrated by the double of $\bar{13}$, probably arises from the Horus-eye representation of the original part.

With this table, which gives the doubles of all parts involving an odd number up to 101, calculations involving parts become feasible. There remains, however, one final complication before one can set out to solve problems. The calculation process described above requires subtraction at each stage in order to find what is lacking in a given column. When the column already contains *parts*, this leads to the second complication: the problem of *subtracting parts*. (*Adding parts* is no problem. The author merely writes them one after another. The sum is condensed if, for example, the author knows that the sum of $\bar{3}$ and $\bar{6}$ is $\bar{2}$.)

This technique, which is harder than the simple procedures discussed above, is explained in the papyrus itself in Problems 21–23. As mentioned above, Problem 21 asks for the parts that must be added to the sum of $\overline{3}$ and $\overline{15}$ to obtain 1. The procedure used to solve this problem is as follows. Begin with the two parts in the first row:

$$\overline{3} \quad \overline{15} \quad 1.$$

Now the problem is to see what must be added to the two terms on the left-hand side in order to obtain the right-hand side. Preserving proportions, the author multiplies the row by 15, getting

$$10 \quad 1 \quad 15$$

It is now clear that when the problem is “magnified” by a factor of 15, we need to add 4 units. Therefore, the only remaining problem is, as we would put it, to divide 4 by 15, or in language that may reflect better the thought process of the author, to “calculate with 15 so as to obtain 4.” This operation is carried out in the usual way:

$$\begin{array}{r} 15 \\ 1 \\ 2 \\ 4 \end{array} \quad \begin{array}{r} 1 \\ \overline{15} \\ \overline{10} \quad \overline{30} \\ \overline{5} \quad \overline{15} \end{array} \quad \begin{array}{l} \\ \text{[from the table]} \\ \text{Result} \end{array}$$

Thus, the parts that must be added to the sum of $\overline{3}$ and $\overline{15}$ in order to reach 1 are $\overline{5}$ and $\overline{15}$. This “subroutine,” which is essential to make the system of computation work, was written in red ink in the manuscripts, as if the writers distinguished between computations made within the problem to find the answer and computations made in order to operate the system. Having learned how to complement (subtract) parts, what are called *hau* (or *aha*) computations by the author, one can confidently attack any arithmetic problem whatsoever. Although there is no single way of doing these problems, specialists in this area have detected (a) systematic procedures by which the table of doubles was generated and (b) patterns in the solution of problems that indicate, if not an algorithmic procedure, at least a certain habitual approach to such problems.

Let us now consider how these principles are used to solve a problem from the papyrus. The one we pick is Problem 35, which, translated literally and misleadingly, reads as follows:

Go down 1 times 3. My third part is added to me. It is filled. What is the quantity saying this?

To clarify: This problem asks for a number that yields 1 when it is tripled and the result is then increased by the third part of the original number. In other words, “calculate with 3 $\overline{3}$ so as to obtain 1.” The solution is as follows:

$$\begin{array}{r} 3 \quad \overline{3} \\ 10 \\ 5 \\ 1 \end{array} \quad \begin{array}{r} 1 \\ 3 \\ 1 \quad \overline{2} \\ \overline{5} \quad \overline{10} \end{array} \quad \begin{array}{l} \\ \text{[multiplied by 3]} \\ \text{Result} \end{array}$$

PROBLEMS AND QUESTIONS

Mathematical Problems

- 6.1. Double the hieroglyphic number $\begin{array}{c} ||| \quad \cap \\ |||| \quad \cap\cap \end{array}$.
- 6.2. Multiply 27 times 42 the Egyptian way.
- 6.3. (Stated in the Egyptian style.) Calculate with 13 so as to obtain 364.

Historical Questions

- 6.4. What are the main documentary sources for our knowledge of ancient Egyptian mathematics?
- 6.5. What benefits are to be gained from learning mathematics, according to the author of the Rhind papyrus?
- 6.6. Is the information in the Rhind papyrus older or more recent than what is found on the Old Babylonian tablets?

Questions for Reflection

- 6.7. Why do you suppose that the author of the Rhind papyrus did not choose to say that the double of the thirteenth part is the seventh part plus the ninety-first part, that is,

$$\frac{2}{13} = \frac{1}{7} + \frac{1}{91}?$$

Why is the relation

$$\frac{2}{13} = \frac{1}{8} + \frac{1}{52} + \frac{1}{104}$$

made the basis for the tabular entry instead?

- 6.8. How do you account for the fact that the ancient Greeks used a system of counting and calculating that mirrored the notation found in Egypt, whereas in their astronomical measurements they borrowed the sexagesimal system of Mesopotamia? Why were they apparently blind to the computational advantages of the place-value system used in Mesopotamia?
- 6.9. Could the ability to solve a problem such as Problem 35 of the Rhind papyrus, discussed above, have been of any practical use? Try to think of a situation in which such a problem might arise.

Algebra and Geometry in Ancient Egypt

Although arithmetic and geometry fill up most of the Egyptian papyri, there are some problems in them that can be considered algebra, provided that we use the very general definition introduced in Chapter 1—that is, the study of techniques for giving the explicit value of a number starting from conditions that determine it implicitly. Most of these problems involve direct proportion and thus lead to linear (first-degree) equations. In the present chapter, we shall examine a selection of these problems and then look at some that have geometric application. In both areas, the goal is to get numerical answers using the computational techniques described in the preceding chapter.

7.1. ALGEBRA PROBLEMS IN THE RHIND PAPYRUS

The concept of proportion is the key to the problems based on the “rule of false position.” Problem 24 of the Rhind papyrus, for example, asks for the quantity that yields 19 when its seventh part is added to it. The author notes that if the quantity were 7 (the “false [sup]position”), it would yield 8 when its seventh part is added to it. Therefore, the correct quantity will be obtained by performing the same operations on the number 7 that yield 19 when performed on the number 8. Thus, we first “calculate with 8 until we reach 19”:

1	8
2	16 *
$\bar{2}$	4
$\bar{4}$	2 *
$\bar{8}$	1 *
2 $\bar{4}$ $\bar{8}$	19 Result

Next, perform these same operations on 7:

1	7
*2	14
$\bar{2}$	3 $\bar{2}$
* $\bar{4}$	1 $\bar{2}$ $\bar{4}$
* $\bar{8}$	$\bar{2}$ $\bar{4}$ $\bar{8}$
2 $\bar{4}$ $\bar{8}$	16 $\bar{2}$ $\bar{8}$ Result

This is the answer. The scribe seems quite confident of the answer and does not carry out the computation needed to verify that it works. Notice that it involves the Horus-eye fractions. These fractions were obviously the easiest to work with, and so occur very frequently in the problems we shall be considering.

The Egyptian scribes were capable of performing operations more complicated than mere proportion. They could take the square root of a number, which they called a *corner*. The Berlin papyrus 6619, contains the following problem (Gillings, 1972, p. 161):

The area of a square of 100 is equal to that of two smaller squares. The side of one is $\bar{2}\bar{4}$ the side of the other. Let me know the sides of the two unknown squares.

Here we are asking for two quantities given their ratio ($\frac{3}{4}$) and the sum of their squares (100). The scribe assumes that one of the squares has side 1 and the other has side $\bar{2}\bar{4}$. Since the resulting total area is $1\bar{2}\bar{16}$, the square root of this quantity is taken ($1\bar{4}$), yielding the side of a square equal to the sum of these two given squares. This side is then multiplied by the correct proportionality factor so as to yield 10 (the square root of 100). That is, the number 10 is divided by $1\bar{4}$, giving 8 as the side of the larger square and hence 6 as the side of the smaller square. This example, incidentally, was cited by van der Waerden as evidence of early knowledge of the Pythagorean theorem in Egypt.

7.1.1. Applied Problems: The *Pesu*

The Rhind papyrus contains problems that involve the concept of proportion in the guise of the slope of pyramids and the strength of beer. Both of these concepts involve what we think of as a ratio, along with the technique of finding the fourth element in a proportion by the procedure once commonly taught to grade-school students and known as the *Rule of Three*. (See Section 2.3 of Chapter 2.) Since the Egyptian procedure for multiplication was based on an implicit notion of proportion, such problems yield easily to the Egyptian techniques, as we shall see below. Several units of weight are mentioned in these problems, but the measurement we shall pay particular attention to is a measure of the dilution of bread or beer. It is called a *pesu* and defined as the number of loaves of bread or jugs of beer obtained from one *hekat* of grain. A hekat was slightly larger than a gallon, 4.8 liters to be precise. Just how much beer or bread it would produce under various circumstances is a technical matter that need not concern us. The thing we need to remember is that the number of loaves of bread or jugs of beer produced by a given amount of grain equals the *pesu* times the number of hekats of grain. A large *pesu* indicates weak beer or bread. In the problems in the Rhind papyrus the *pesu* of beer varies from 1 to 4, while that for bread varies from 5 to 45.

Problem 71 tells of a jug of beer produced from half a hekat of grain (thus its *pesu* was 2). One-fourth of the beer is poured off, and the jug is topped up with water. The problem asks for the new *pesu*. The author reasons that the eighth part of a hekat of grain was removed, leaving (in his terms) $\bar{4}\bar{8}$, that is, what we would call $\frac{3}{8}$ of a hekat of grain. Since this amount of grain goes into one jug, it follows that the *pesu* of that beer is what we call the *reciprocal* of that number, namely $2\bar{3}$. The author gives this result immediately, apparently assuming that by now the reader will know how to “calculate with $\bar{4}\bar{8}$ until 1 is reached.”

The Rule of Three procedure is invoked in Problem 73, which asks how many loaves of 15-*pesu* bread are required to provide the same amount of grain as 100 loaves of 10-*pesu*

bread. The answer is found by dividing 100 by 10, then multiplying by 15, which is precisely the Rule of Three.

7.2. GEOMETRY

The most fascinating aspect of Egyptian mathematics is the application of these computational techniques to geometry. In Section 109 of Book 2 of his *History*, the Greek historian Herodotus writes that King Sesostris¹ dug a multitude of canals to carry water to the arid parts of Egypt. He goes on to connect this Egyptian engineering with Greek geometry:

It was also said that this king distributed the land to all the Egyptians, giving an equal quadrilateral farm to each, and that he got his revenue from this, establishing a tax to be paid for it. If the river carried off part of someone's farm, that person would come and let him know what had happened. He would send surveyors to remeasure and determine the amount by which the land had decreased, so that the person would pay less tax in proportion to the loss. It seems likely to me that it was from this source that geometry was found to have come into Greece. For the Greeks learned of the sundial and the twelve parts of the day from the Babylonians.

The main work of Egyptian surveyors was measuring fields. That job is literally described by its Latin name *agrimensor*. Our word *surveyor* comes through French, but has its origin in the Latin *supervideo*, meaning *I oversee*. The equivalent word in Greek was used by Herodotus in the passage above. He described *episkepsoménous kai anametrēsontas* (ἐπισκεψομένους καὶ ἀναμετρήσοντας), using future participles that mean literally “[people who] will be inspecting for themselves and measuring carefully.”² The process of measuring a field is shown in a painting from the tomb of an Egyptian noble named Menna at Sheikh Abd el-Qurna in Thebes. Menna bore the title Scribe of the Fields of the Lord of the Two Lands during the Eighteenth Dynasty, probably in the reign of Amenhotep III or Thutmose IV, around 1400 BCE. His job was probably that of a steward, to oversee planting and harvest. The instrument used to measure distance was a rope that could be pulled taut. That measuring instrument has given rise to another name often used to refer to these surveyors: *harpedonáptai*, from the words *harpedónē*, meaning *rope*, and *háptō*, meaning *I attach*. The philosopher Democritus (d. 357 BCE) boasted, “In demonstration no one ever surpassed me, not even those of the Egyptians called *harpedonáptai*.”³

The geometric problems considered in the Egyptian papyri all involve numerical measurement rather than the more abstract proportions that make up the bulk of Greek geometry. These problems show considerable insight into the properties of simple geometric figures such as the circle, the triangle, the rectangle, and the pyramid; and they rise to a rather high level of sophistication in computing the area of a hemisphere. The procedures for measuring regions with flat boundaries (polygons and pyramids) are correct from the point of view of Euclidean geometry, while those involving curved boundaries (disks and spheres) have an error controlled entirely by the error in the ratio the area of a circle bears to the square on

¹There were several pharaohs with this name. Some authorities believe that the one mentioned by Herodotus was actually Ramses II, who ruled from 1279 to 1212 BCE.

²Or “remeasuring.”

³Quoted by the second-century theologian Clement of Alexandria, in his *Stromata (Miscellanies)*, Book 1, Chapter 15.

its diameter. In the papyrus, this ratio is given as $(\frac{8}{9})^2 \approx 0.79012345679 \dots$. In Euclidean geometry, it is $\frac{\pi}{4} \approx 0.785398163397 \dots$.

7.3. AREAS

Since the areas of rectangles and triangles are easy to compute, it is understandable that very little attention is given to these problems. Only four problems in the Rhind papyrus touch on these questions, namely Problems 6, 49, 51, and 52.

7.3.1. Rectangles, Triangles, and Trapezoids

Problem 49 involves computing the area of a rectangle that has dimensions 1 *khet* by 10 *khet*s. This in itself would be a trivial problem, except that areas are to be expressed in square cubits rather than square *khet*s. Since a *khet* is 100 cubits, the answer is given correctly as 100,000 square cubits. Problem 51 is a matter of finding the area of a triangle, and it is illustrated by a figure showing the triangle. The area is found by multiplying half of the base by the height. In Problem 52, this technique is generalized to a trapezoid, and half of the sum of the upper and lower bases is multiplied by the height.

Of all these problems, the most interesting is Problem 6, which involves a twist that makes it equivalent to a quadratic equation. A rectangle is given having area 12 *cubit strips*; that is, it is equal to an area 1 cubit by 12 cubits, though not of the same shape. The problem is to find its dimensions given that the width is three-fourths of the length ($\bar{2} \bar{4}$ in the notation of the papyrus). The first problem is to “calculate with $\bar{2} \bar{4}$, until 1 is reached,” that is, in our language, dividing 1 by $\bar{2} \bar{4}$. The result is $1 \bar{3}$. Then 12 is multiplied by $1 \bar{3}$, yielding 16, after which the scribe takes the *corner* (square root) of 16, getting 4 as the length. This is a very nice example of thinking in terms of descriptions that determine a quantity and using those descriptions to exhibit its value explicitly, exactly what we are defining algebra to be. The scribe seems to have in mind a picture of the length being multiplied by three-fourths of the length, the result being 12. This 12, which is $\frac{3}{4}$ of the square of the length, is multiplied by the reciprocal of $\frac{3}{4}$, after which the length is found by taking the square root. From the scribe’s point of view, the heart of the problem was getting the reciprocal of $\frac{3}{4}$.

7.3.2. Slopes

Given that the Egyptians had no trigonometry as we now understand it, it is interesting to observe the solutions of problems that involve the slope of the sides of pyramids and other figures. There is a unit of slope analogous to the *pesu* that we have just seen in the problems involving strength of bread and beer. The unit of slope is the *seked*, defined as the number of palms of horizontal displacement associated with a vertical displacement of 1 royal cubit. One royal cubit was 7 palms. Because of the relative sizes of horizontal and vertical displacements, it makes sense to use the larger unit of length (the cubit) for vertical distances and the smaller one (the palm) for horizontal distances, even at the expense of introducing an extra factor into computations of slope. In our terms the *seked* is seven times the tangent of the angle that the sloping side makes with the vertical. In some of the problems the *seked* is given in such a way that the factor of 7 drops out. Notice that if you were ordering a stone from the quarry, the *seked* would tell the stonemason immediately where to cut. One would

mark a point one cubit (distance from fingertip to elbow) from the corner in one direction and a point at a number of palms equal to the *seked* in the perpendicular direction and then simply cut between the two points marked.

In Problem 57 a pyramid with a *seked* of $5\bar{4}$ and a base of 140 cubits is given. The problem is to find its height. The *seked* given here ($\frac{3}{4}$ of 7) is exactly that of one of the actual pyramids, the pyramid of Khafre, who reigned from 2558 to 2532 BCE. It appears that stones were mass-produced in several standard shapes with a *seked* that could be increased in intervals of one-fourth. Pyramid builders and designers could thereby refer to a standard brick shape, just as architects and contractors since the time of ancient Rome have been able to specify a standard diameter for a water pipe. Problem 58 gives the dimensions of the same pyramid and asks for its *seked*, apparently just to reinforce the reader's grasp of the relation between *seked* and dimension.

7.3.3. Circles

Five of the problems in the Rhind papyrus (41–43, 48, and 50) involve calculating the area of a circle. The answers given are approximations, but, as mentioned above, would be precise if the value $64/81$ used in the papyrus where we would use $\pi/4$ were exact. The author makes no comment suggesting that this value is only an approximation. Nor should we expect him to, since he would have had no concept of infinite precision in measuring continuous objects.

A Digression: Commercial Computation of Volumes. When physical objects such as grain silos are built, the parts used to build them have to be measured. In addition, the structures and their contents have a commercial, monetary value. Some number has to be used to express that value. It would therefore *not* be absurd—although it would probably be unnecessary—for a legislature to pass a bill prescribing a numerical value to be used for π .⁴ Similarly, the claim often made that the “biblical” value of π is 3, based on the description of a vat 10 cubits from brim to brim girdled by a line of 30 cubits (1 Kings 7:23) is pure pedantry. It assumes more precision than is necessary in the context. The author may have been giving measurements only to the nearest 10 cubits, not an unreasonable thing to do in a literary description.⁵ We now return to the subject of circle measurements in Egypt.

⁴However, in the most notorious case where such a bill was nearly passed—House Bill 246 of the 1897 Indiana legislature—it *was* absurd. The bill was written by a physician and amateur mathematician named Edwin J. Goodwin. Goodwin had copyrighted what he thought was a quadrature of the circle. He offered to allow textbooks sold in Indiana to use his proof royalty-free provided that the Indiana House would pass this bill, whose text mostly glorified his own genius. Some of the mathematical statements the legislature was requested to enact were pure gibberish. For example, “a circular area is to the square on a line equal to the quadrant of the circumference, as the area of an equilateral rectangle is to the square on one side.” The one clear statement is that “the ratio of the chord and arc of ninety degrees. . . is as seven to eight.” That statement implies that $\pi = 16\sqrt{2}/7 \approx 3.232488 \dots$. The square root in this expression did not trouble Dr. Goodwin, who declared that $\sqrt{2} = 10/7$. At this point, one might have taken his value of π to be $160/49 = 3.265306122 \dots$. But, in a rare and uncalled-for manifestation of consistency, since he “knew” that $100/49 = (10/7)^2 = 2$, Goodwin declared this fraction equal to $16/5 = 3.2$. The bill was stopped at the last minute by lobbying from a member of the Indiana Academy of Sciences and was tabled without action.

⁵However, like everything in the Bible, this passage has been subject to repeated analysis. For a summary of the conclusions reached in the Talmud, see Tsaban and Garber (1998).

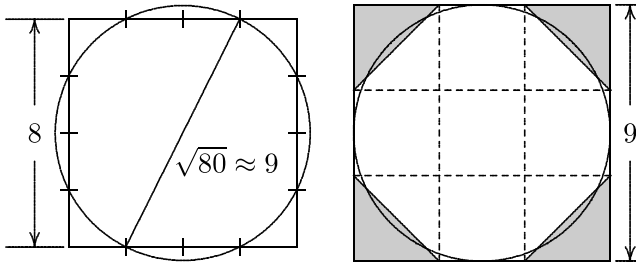


Figure 7.1. Conjectured explanations of the Egyptian squaring of the circle.

Ahmosé takes the area of a circle to be the area of the square whose side is obtained by removing the ninth part of the diameter. In our language the area is the square on eight-ninths of the diameter; that is, it is the square on $\frac{16}{9}$ of the radius. In our language, not that of Egypt, this gives a value of π for area problems equal to $\frac{256}{81}$. Please remember, however, that the Egyptians had no concept of the number π . The constant of proportionality that they always worked with represents what we would call $\pi/4$. There have been various conjectures as to how the Egyptians might have arrived at this result. One such conjecture, given by Robins and Shute (1987, p. 45), involves a square of side 8. If a circle is drawn through the points 2 units from each corner, it is visually clear that the four fillets at the corners, at which the square is outside the circle, are nearly the same size as the four segments of the circle outside the square; hence this circle and this square may be considered equal in area. Now the diameter of this circle can be obtained by connecting one of the points of intersection to the opposite point, as shown on the left-hand diagram in Fig. 7.3, and measurement will show that this line is very nearly 9 units in length (it is actually $\sqrt{80}$ in length). A second theory due to K. Vogel [see Gillings (1972, pp. 143–144)] is based on the fact that the circle inscribed in a square of side nine is roughly equal to the unshaded region in the right-hand diagram in Fig. 7.1. This area is $\frac{7}{9}$ of 81, that is, 63. A square of equal size would therefore have side $\sqrt{63} \approx 7.937 \approx 8$. In favor of Vogel's conjecture is the fact that a figure very similar to this diagram accompanies Problem 48 of the papyrus. A discussion of various conjectures, giving connections with traditional African crafts, was given by Gerdes (1985).

7.3.4. The Pythagorean Theorem

In the discussion of ancient cultures, the question of the role played by the Pythagorean theorem is of interest. Did the ancient Egyptians know this theorem? It has been reported in numerous textbooks, popular articles, and educational videos that the Egyptians laid out right angles by stretching a rope with 12 equal intervals knotted on it so as to form a 3–4–5 right triangle. What is the evidence for this assertion? First, the Egyptians *did* lay out very accurate right angles. Also, as mentioned above, it is known that their surveyors used ropes as measuring instruments and were referred to as *rope-fixers*. That is the evidence that was cited by the person who originally made the conjecture, the historian Moritz Cantor (1829–1920) in the first volume of his history of mathematics, published in 1880. The case can be made stronger, however. In his essay *Isis and Osiris*, the first-century polymath Plutarch says the following.

It has been imagined that the Egyptians regarded one triangle above all others, likening it to the nature of the universe. And in his *Republic* Plato seems to have used it in arranging marriages.

This triangle has 3 on the vertical side, 4 on the base, and a hypotenuse of 5, equal in square to the other two sides. It is to be imagined then that it was constituted of the masculine on the vertical side, and the feminine on the base; also, Osiris as the progenitor, Isis as the receptacle, and Horus as the offspring. For 3 is the first odd number and is a perfect number;⁶ the 4 is a square formed from an even number of dyads; and the 5 is regarded as derived in one way from the father and another way from the mother, being made up of the triad and the dyad.

Finally, as mentioned above, Berlin papyrus 6619 contains a problem in which one square equals the sum of two others. It is hard to imagine anyone being interested in such conditions without knowing the Pythagorean theorem. Against the conjecture, we could note that the earliest Egyptian text that mentions a right triangle and finds the length of all its sides using the Pythagorean theorem dates from about 300 BCE, and by that time the presence of Greek mathematics in Alexandria was already established. None of the older papyri mention or use by implication the Pythagorean theorem.

On balance, one would guess that the Egyptians *did* know the Pythagorean theorem. However, there is no evidence that they used it to construct right angles, as Cantor conjectured. There are much simpler ways of doing that (even involving the stretching of ropes), which the Egyptians must have known. Given that the evidence for this conjecture is so meager, why is it so often reported as fact? Simply because it has been repeated frequently since it was originally made. We know precisely the source of the conjecture, but that knowledge does not seem to reach the many people who report it as fact.⁷

7.3.5. Spheres or Cylinders?

Problem 10 of the Moscow papyrus has been subject to various interpretations. It asks for the area of a curved surface that is either half of a cylinder or a hemisphere. In either case it is worth noting that the area is obtained by multiplying the length of a semicircle by another length in order to obtain the area. Finding the area of a hemisphere is an extremely difficult problem. Intuitive techniques that work on flat or ruled surfaces break down. If the Egyptians did compute this area, no one has given any reasonable conjecture as to how they did so. The difficulty of this problem was given as one reason for interpreting the figure as half of a cylinder. Yet the plain language of the problem implies that the surface is a hemisphere. The problem was translated into German by the Russian scholar V. V. Struve (1889–1965); the following is a translation from the German:

The way of calculating a basket, if you are given a basket with an opening of $4\bar{2}$. O, tell me its surface!

⁶The number 3 is not perfect according to the Euclidean definition, as Plutarch must have known. He uses the same words that Euclid uses for odd number (*perissós*, meaning a number having an excess (when divided by 2), and perfect number (*téleios*). Euclid defines a number to be perfect if it is equal to [the sum of] the numbers that measure it, that is, divide it evenly. If 1 is counted as a number that measures it and the number itself is not, then no prime can be perfect. If the opposite convention is adopted (since the Greeks generally didn't think of 1 as a number), then primes and perfect numbers are the same thing. What could Plutarch have been thinking?

⁷This point was made very forcefully by van der Waerden (1963, p. 6). In a later book (1983), van der Waerden claimed that integer-sided right triangles, which seem to imply knowledge of the Pythagorean theorem, are ubiquitous in the oldest megalithic structures. Thus, he seems to imply that the Egyptians knew the theorem, but didn't use it as Cantor suggested.

Calculate $\frac{1}{9}$ of 9, since the basket is half of an egg. The result is 1. Calculate what is left as 8. Calculate $\frac{1}{9}$ of 8. The result is $\frac{8}{9}$. Calculate what is left of this 8 after this $\frac{8}{9}$ is taken away. The result is $\frac{72}{9}$. Calculate $\frac{1}{2}$ times with $\frac{72}{9}$. The result is 32. Behold, this is the surface. You have found it correctly.

If we interpret the basket as being a hemisphere, the scribe has first doubled the diameter of the opening from $\frac{1}{2}$ to 9 “because the basket is half of an egg.” (If it had been the *whole* egg, the diameter would have been quadrupled.) The procedure used for finding the area here is equivalent to the formula $2d \cdot \frac{8}{9} \cdot \frac{8}{9} \cdot d$. Taking $(\frac{8}{9})^2$ as representing $\pi/4$, we find it equal to $(\pi d^2)/2$, or $2\pi r^2$, which is indeed the area of a hemisphere of radius r .

Van der Waerden points out (1963, pp. 33–34) that this value is also the lateral area of half of a cylinder of height d and base diameter d if it is laid on its side and bisected by a plane through the diameters of its two circular bases. In that case, the two bases are not counted as part of the area, and the basket must be regarded as a semicircular lamina attached to the two opposite sides of its top.⁸ In this case, the opening would be square, since its width is the height of the cylinder, its length is the diameter of the base, and the two are equal; the number $\frac{1}{2}$ would be the side of the square. That would mean also that the “Egyptian π ” ($\pi/4 = 64/81$), used for area problems, which we refer to as *two-dimensional* π , was also being applied to the ratio of the circumference to the diameter, which we refer to as *one-dimensional* π . In other words, the Egyptians would have known that the ratio of the circumference to the diameter of a circle is the same as the ratio of the square on its radius to the disk it encloses. The numerical answer is consistent with this interpretation, but, as just mentioned, only the lateral surface of the cylinder is to be included. That would indicate that the basket was open at the sides. It would be strange to describe such a basket as “half of an egg.” The main reason given by van der Waerden for preferring this interpretation is an apparent inaccuracy in Struve’s statement of the problem. Van der Waerden quotes T. E. Peet, who says that the number $\frac{1}{2}$ occurs twice in the statement of the problem, as the opening of the top of the basket and also as its depth. This interpretation, however, leads to further difficulties. If the surface is indeed half of a cylinder of base diameter $\frac{1}{2}$, its depth is not $\frac{1}{2}$; it is $\frac{1}{4}$. Van der Waerden also mentions a conjecture of Neugebauer, that this surface was intended to be a domelike structure of a sort seen in some Egyptian paintings, resembling very much the small end of an egg. That interpretation restores the idea that this problem was the computation of the area of a nonruled surface, and the approximation just happens to be the area of a hemisphere.

7.3.6. Volumes

One of the most remarkable achievements of the Egyptians is the discovery of accurate ways of computing volumes. In Problem 41 of the Rhind papyrus we find the correct procedure used for finding the volume of a cylindrical silo, that is, the area of the circular base is multiplied by the height. To make the numbers easy, the diameter of the base is given as 9 cubits, as in Problems 48 and 50, so that the area is 64 square cubits. The height is 10 cubits, giving a volume of 640 cubic cubits. However, the standard unit of grain volume was a *khar*, which is two-thirds of a cubic cubit, resulting in a volume of 960 *khar*. In a further

⁸If the cylinder is truncated by a plane parallel to its base and at height equal to half of the radius of the base, then one also gets the correct area. That is, the basket is an upright cylinder whose height is half the radius of its base.

twist, to get a smaller answer, the scribe divides this number by 20, getting 48 “hundreds of quadruple *hekats*.” (A *khar* was 20 quadruple *hekats*.) Problem 42 is the same problem, only with a base of diameter 10 cubits. Apparently, once the reader has the rule well in hand, it is time to test the limits by making the data more cumbersome. The answer is computed to be $1185 \bar{6} \bar{54}$ *khar*, again expressed in hundreds of quadruple *hekats*. Problems 44–46 calculate the volume of prisms on a rectangular base by the same procedure.

Given that pyramids are so common in Egypt, it is surprising that the Rhind papyrus does not discuss the volume of a pyramid. However, Problem 14 from the Moscow papyrus asks for the volume of the frustum of a square pyramid, given that the side of the lower base is 4, the side of the upper base is 2, and the height is 6. The author gives the correct procedure: Add the areas of the two bases to the area of the rectangle whose sides are the sides of the bases, that is, $2 \cdot 2 + 4 \cdot 4 + 2 \cdot 4$, then multiply by one-third of the height, getting 56. This technique could not have been arrived at through experience. Some geometric principle must be involved, since the writer knew that the sides of the bases, which are *parallel* lines, need to be multiplied. Normally, the lengths of two lines are multiplied only when they are perpendicular to each other, so that the product represents the area of a rectangle. Gillings (1972, pp. 190–193) suggests a possible route by which this knowledge may have been obtained. Robins and Shute (1987, pp. 48–49) suggest that the result may have been obtained by completing the frustum to a full pyramid and then subtracting the volume of the smaller pyramid from the larger. In either case, the power of visualization involved in seeing that the procedure will work is remarkable.

Like the surface area problem from the Moscow papyrus just discussed, this problem reflects a level of geometric insight that must have required some accumulation of observations built up over time. It is very easy to see that if a right pyramid with a square base is sliced in half by a plane through its vertex and a pair of diagonally opposite vertices of the base, the base is bisected along with the pyramid. Thus, a tetrahedron whose base is half of a square has volume exactly half that of the pyramid of the same height having the whole square as a base.

It is also easy to visualize how a cube can be cut into two wedges, as in the top row of Fig. 7.2. Each of these wedges can then be cut into a pyramid on a face of the cube plus an extra tetrahedron, as in the bottom row. The tetrahedron $P'Q'R'S'$ has a base $P'Q'R'$ that is half of the square base $PQRT$ of the pyramid $PQRST$ and hence has half of its volume. It follows that the volume of the tetrahedron is one-sixth that of the cube, and so the pyramid $PQRST$ is one-third of the volume. A “mixed” geometric-mechanical strategy is also possible, involving weighing of the parts. The two tetrahedra would, in theory, balance one of the square pyramids. This model could be sawn out of stone or wood. From that special case one might generalize the vital clue that the volume of a pyramid is one-third the area of the base times the altitude.

Once the principle is established that a pyramid equals a prism on the same base with one-third the height, it is not difficult to chop a frustum of a pyramid into the three pieces described in the Moscow papyrus. Referring to Fig. 7.3, which shows a frustum with bottom base a square of side a and upper base a square of side b with $b < a$, we can cut off the four corners and replace them by four rectangular solids with square base of side $(a - b)/2$ and height $h/3$. These four fit together to make a single solid with square base of side $a - b$ and height $h/3$. One opposite pair of the four sloping faces that remain after the corners are removed can be cut off, turned upside down, and laid against the other two sloping faces so as to form a rectangular prism with a rectangular base that is $a \times b$ and has height h . The top one-third of this prism can then be cut off and laid aside. It has volume $(h/3)ab$. The

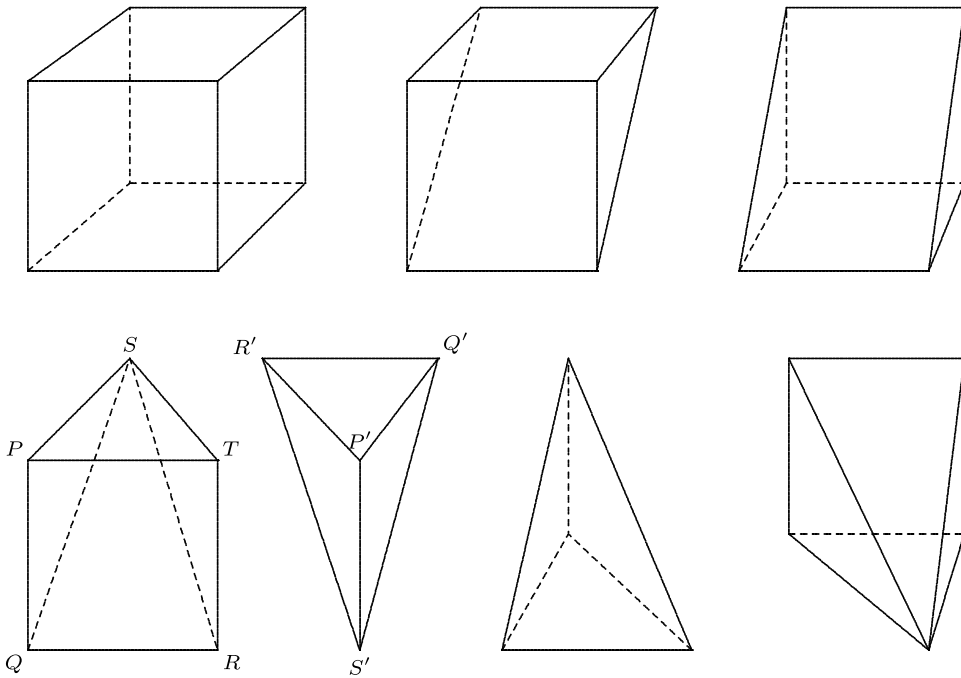


Figure 7.2. Dissection of a cube into two square pyramids and two tetrahedra.

top half of what remains can then be cut off, and a square prism of base side b and height $h/3$ can be cut off from it. If that square prism is laid aside (it has volume $(h/3)b^2$), the remaining piece, which is $(a - b) \times b \times (h/3)$, will fill out the other corner of the bottom layer, resulting in a square prism of volume $(h/3)a^2$. Thus, we obtain the three pieces that the scribe added to get the volume of the frustum in a way that is not terribly implausible.

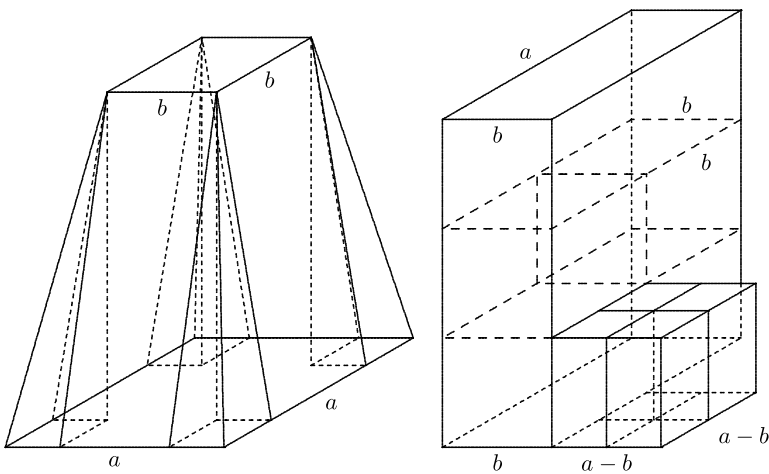


Figure 7.3. Dissection of a frustum of a pyramid.

These last few paragraphs and Figs. 7.2 and 7.3 are conjectures, not facts of history. We do not know how the Egyptians discovered that the volume of a pyramid is one-third the volume of a prism of the same base and height or how they learned how to compute the volume of a frustum.

PROBLEMS AND QUESTIONS

Mathematical Problems

- 7.1. Compare the *pesu* problems in the Rhind papyrus with the following problem, which might have been taken from almost any algebra book written in the past century: *A radiator is filled with 16 quarts of a 10% alcohol solution. If it requires a 30% alcohol solution to protect the radiator from freezing, how much 95% solution must be added (after an equal amount of the 10% solution is drained off) to provide this protection? Think of the alcohol as the grain in beer and the liquid in the radiator as the beer. The liquid has a *pesu* of 10. What is the *pesu* that it needs to have, and what is the *pesu* of the liquid that is to be used to achieve this result?*
- 7.2. Problem 33 of the Rhind papyrus asks for a quantity that yields 37 when increased by its two parts (two-thirds), its half, and its seventh part. Try to get the author's answer: The quantity is $16 \frac{56}{679} \frac{776}{776}$. [*Hint*: The table for doubling fractions gives the last three terms of this expression as the double of $\frac{97}{97}$. The scribe first tried the number 16 and found that the result of these operations applied to 16 fell short of 37 by the double of $\frac{42}{42}$, which, as it happens, is exactly $1 \frac{3}{3} \frac{2}{2} \frac{7}{7}$ times the double of $\frac{97}{97}$.]
- 7.3. Find the height of the pyramid with a square base of side 140 cubits and *seked* equal to $5 \frac{4}{4}$ (Problem 57 of the Rhind papyrus).

Historical Questions

- 7.4. Compare the procedures for computing volumes in ancient Mesopotamia with those used in ancient Egypt.
- 7.5. To what audience does the Rhind papyrus appear to be addressed?
- 7.6. What principles seem to determine the order of the problems discussed in the Rhind papyrus?

Questions for Reflection

- 7.7. Why not simply write $\frac{13}{13} \frac{13}{13}$ to stand for what we call $\frac{2}{13}$? What is the reason for using two or three other "parts" instead of these two obvious parts?
- 7.8. We would naturally solve many of the problems in the Rhind papyrus using an equation. Would it be appropriate to say that the Egyptians solved equations, or that they did algebra?
- 7.9. What do you imagine was the social position of Ahmose, who wrote the Rhind papyrus? What were his normal duties, and for what purpose did he undertake this labor?

GREEK MATHEMATICS FROM 500 BCE TO 500 CE

During the millennium from 500 BCE to 500 CE, mathematics, especially geometry, was imported into Greece, became mixed with the speculations of the Greek philosophers and developed into a body of knowledge that was unique in its time. The center of gravity gradually shifted from the commercial Ionian colonies in the early period (along with their colonies in Italy and Sicily) to Athens in the fifth century, reaching its peak in the third century at Alexandria, Egypt. Along with this new, formal geometry, some of the earlier, more practically oriented geometry survived and revived after the time of the three greatest geometers of antiquity (Euclid, Archimedes, and Apollonius). This practical geometry was applied to produce one great astronomical treatise, the *Almagest* of Claudius Ptolemy, which became a standard reference for the next thousand years. After Ptolemy, the potential of the Euclidean methods was nearly exhausted, and there is little that is original in the last few centuries we are going to discuss. The one exception comes from number theory, rather than geometry, and it is the invention of symbolic algebraic notation by Diophantus in order to solve problems involving the arithmetic properties of figurate numbers. The next 11 chapters give a sketch of some of the highlights of this long period of development.

Contents of Part III

1. Chapter 8 (An Overview of Ancient Greek Mathematics) gives a survey of the whole period and summarizes the sources on which our knowledge of it is based.
2. Chapter 9 (Greek Number Theory) looks at ancient Greek number theory through the works of Euclid, Nicomachus, and Diophantus.
3. Chapter 10 (Fifth-Century Greek Geometry) presents a hypothetical scenario for the development of geometry up to the mid-fifth century BCE.
4. Chapter 11 (Athenian Mathematics I: The Classical Problems) brings the development of geometry in Athens to the end of the fourth century BCE.
5. Chapter 12 (Athenian Mathematics II: Plato and Aristotle) discusses the connection of this geometry with the philosophies of Plato and Aristotle.
6. Chapter 13 (Euclid of Alexandria) analyzes the *Elements* and looks briefly at some other works by Euclid.

7. Chapter 14 (Archimedes of Syracuse) is devoted to the works of Archimedes.
8. Chapter 15 (Apollonius of Perga) discusses the extensive treatise on conic sections by Apollonius of Perga.
9. Chapter 16 (Hellenistic and Roman Geometry) discusses the isoperimetric problems studied by Zenodorus and the return of metric concepts to geometry in the work of Heron of Alexandria.
10. Chapter 17 (Ptolemy's Geography and Astronomy) is devoted to the geographical and astronomical treatises of Claudius Ptolemy.
11. Chapter 18 (Pappus and the Later Commentators) summarizes the work of the later commentators Pappus, Theon of Alexandria, and Theon's daughter Hypatia.

An Overview of Ancient Greek Mathematics

Greek was the common language of scholarship in the region around the Mediterranean for at least nine hundred years, from the time of the Athenian Empire in the mid-fifth century BCE until the Western half of the Roman Empire was destroyed by invaders from the north in the late fifth century CE, after which contacts between the Western Latin-based portion and the Eastern Greek-speaking portion began to decline. The Roman Empire was multinational, and not all those who contributed to this scholarship were native speakers of Greek. Although the great classic works of the third century BCE probably were written by native Greek speakers—Archimedes certainly was one—some of the later commentators may have had other roots. What we call ancient Greek mathematics is therefore mathematics originally written *in* Greek, not necessarily *by* Greeks, and some of it survives only in Arabic translation.

The origin, flourishing, and decline of ancient Greek mathematics took place over a period approximately 1000 years in extent, beginning with the philosopher Thales (ca. 624–546) and ending with the death of Hypatia in 415 CE. It arose during the Hellenic era from 600 to 300 BCE, when the Greek city-states were independent, achieved its greatest heights during the Hellenistic period after the conquests of Alexander the Great, and underwent stagnation and decline after the rise of the Roman Empire, while still producing some remarkable works during its final 500 years.

The Greeks of the Hellenic period traced the origins of their mathematical knowledge to Egypt and the Middle East. This knowledge probably came in “applied” form in connection with commerce and astronomy/astrology. Mesopotamian numerical methods appear in the later Hellenistic work on astronomy by Hipparchus (second century BCE) and the work of Ptolemy under the Roman Empire. Earlier astronomical models by Eudoxus (fourth century BCE) and Apollonius (third century BCE) were more geometrical. Jones (1991, p. 445) notes that “the astronomy that the Hellenistic Greeks received from the hands of the Babylonians was by then more a skill than a science: The quality of the predictions was proverbial, but in all likelihood the practitioners knew little or nothing of the origins of their schemes in theory and observations.” Among the techniques transmitted to the Greeks and ultimately to the modern world was the convention of dividing a circle into 360 equal parts (degrees). Greek astronomers divided the radius into 60 equal parts so that the units of length on the radius and on the circle were very nearly equal.

1	2	3	4	5	6	7	8	9
α'	β'	γ'	δ'	ε'	F'	ζ'	η'	θ'
10	20	30	40	50	60	70	80	90
ι'	κ'	λ'	μ'	ν'	ξ'	\omicron'	π'	ϕ'
100	200	300	400	500	600	700	800	900
ρ'	σ'	τ'	υ'	φ'	χ'	ψ'	ω'	λ'

Figure 8.1. The ancient Greek numbering system.

The amount that the Greeks learned from Egypt is the subject of controversy. Some scholars who have read the surviving mathematical texts from papyri have concluded that Egyptian methods of computing were too cumbersome for application to the complicated measurements of astronomers. Yet both Plato and Aristotle speak approvingly of Egyptian computational methods and the ways in which they were taught. As for geometry, it is generally acknowledged that the Egyptian insight was extraordinary; the Egyptians knew how to find the volume of a frustum of a pyramid, and it appears that they even found the area of a hemisphere, the only case known before Archimedes in which the area of a curved surface is found.¹ The case for advanced Egyptian mathematics is argued by Bernal (1992), who asserts that Ptolemy himself was an Egyptian. The question is difficult to settle, since little is known of Ptolemy personally; for us, he is simply the author of certain works on physics, astronomy, and geography. One particular aspect of Greek mathematics, however, does bear a strong resemblance to that of Egypt, namely their system of writing numbers. It is shown here in Fig. 8.1, which is to be compared with Fig. 6.2 of Chapter 6.

Because of their extensive commerce, with its need for counting, measurement, navigation, and an accurate calendar, the Ionian Greek colonies such as Miletus on the coast of Asia Minor and Samos in the Aegean Sea provided a very favorable environment for the development of mathematics, and it was there, with the philosophers Thales of Miletus and Pythagoras of Samos (ca. 570–475 BCE), that Greek mathematics began.

8.1. SOURCES

Since the material on which the Greeks wrote was not durable, all the original manuscripts have been lost except for a few ostraca (shells) found in Egypt. We are dependent on copyists for preserving the information in early Greek works, since few manuscripts that still exist were written more than 1000 years ago. We are further indebted to the many commentators who wrote summary histories of philosophy, including mathematics, for the little that we know about the works that have not been preserved and their authors. The most prominent among these commentators are listed below. They will be mentioned many times in the chapters that follow.

1. Marcus Vitruvius (first century BCE) was a Roman architect who wrote a treatise on architecture in 10 books. He is regarded as a rather unreliable source for information about mathematics, however.

¹As mentioned in the preceding chapter, some authors claim that the surface in question was actually half of a cylinder, but the words used seem more consistent with a hemisphere. In either case it was a curved surface.

2. Plutarch (45–120 CE) was a pagan author, apparently one of the best educated people of his time, who wrote on many subjects. He is best remembered as the author of the *Parallel Lives of the Greeks and Romans*, in which he compares famous Greeks with eminent Romans who engaged in the same occupation, such as the orators Demosthenes and Cicero.² Plutarch is important to the history of mathematics for what he reports on natural philosophers such as Thales.
3. Theon of Smyrna (ca. 100 CE) was the author of an introduction to mathematics written as background for reading Plato, a copy of which still exists. It contains many quotations from earlier authors.
4. Diogenes Laertius (third century CE) wrote a comprehensive history of philosophy, *Lives of Eminent Philosophers*, which contains summaries of many earlier works and gives details of the lives and work of many of the pre-Socratic philosophers. He appears to be the source of the misnomer “Pythagorean theorem” that has come down to us (see Zhmud, 1989, p. 257).
5. Iamblichus (285–330 CE) was the author of many treatises, including 10 books on the Pythagoreans, five of which have been preserved.
6. Pappus (ca. 300 CE) wrote many books on geometry, including a comprehensive treatise of eight mathematical books. He is immortalized in calculus books for his theorem on the volume of a solid of revolution. Besides being a first-rate geometer in his own right, he wrote commentaries on the *Almagest* of Ptolemy and the tenth book of Euclid’s *Elements*.
7. Theon of Alexandria (late fourth century CE), a commentator and philosopher who is probably responsible for the now-standard Greek edition of Euclid’s *Elements*.
8. Hypatia of Alexandria (ca. 370–415), a neo-Platonist philosopher, daughter of Theon of Alexandria. She may be the editor of the Greek text of Diophantus’ *Arithmetica*.
9. Proclus (412–485 CE) is the author of a commentary on the first book of Euclid, in which he seems to have quoted a long passage from a history of mathematics, now lost, by Eudemus, a pupil of Aristotle.
10. Simplicius (500–549 CE) was a commentator on philosophy. His works contain many quotations from the pre-Socratic philosophers.
11. Eutocius (ca. 700 CE) was a mathematician who lived in the port city of Askelon in Palestine and wrote an extensive commentary on the works of Archimedes.

8.1.1. Loss and Recovery

Most of these commentators wrote in Greek. Knowledge of Greek sank to a very low level in western Europe as a result of the upheavals of the fifth century. Although learning was preserved by the Catholic Church and all of the New Testament was written in Greek, a Latin translation (the Vulgate) was made by Jerome in the fifth century. From that time on, Greek documents were preserved mostly in the Eastern (Byzantine) Empire. After the Muslim conquest of North Africa and Spain in the eighth century, some Greek documents were translated into Arabic and circulated in Spain and the Middle East. From the eleventh century on, as secular learning began to revive in the West, scholars from northern Europe

²Shakespeare relied on Plutarch’s account of the life of Julius Caesar, even describing the miraculous omens that Plutarch reported as having occurred just before Caesar’s death.

made journeys to these centers and to Constantinople, copied out manuscripts, translated them from Arabic and Greek into Latin, and tried to piece together some long-forgotten parts of ancient learning.

8.2. GENERAL FEATURES OF GREEK MATHEMATICS

Greek mathematics—that is, mathematics written in ancient Greek—is exceedingly rich in authors and works. Its most unusual feature, compared with what went before, is its formal development. From the time of Euclid on, mathematics was developed systematically from definitions and axioms, general theorems were stated, and proofs were given. This formal development is the outcome of the entanglement of mathematics with Greek philosophy. It became a model to be imitated in many later scientific treatises, such as Newton's *Philosophiæ naturalis principia mathematica*. Of course, Greek mathematics did not arise in the finished form found in the treatises. Tradition credits Thales, the earliest Greek philosopher, with knowing four geometric propositions. Thales was said to have traveled to Egypt and determined the height of the Great Pyramid of Khufu using similar triangles. One of the four geometric propositions that Thales is said to have known is that an angle inscribed in a semicircle is a right angle.³

Herodotus mentions Thales in several places. Discussing the war between the Medes and the Lydian king Croesus, which had taken place in the previous century, he says that an eclipse of the sun frightened the combatants into making peace. Thales, according to Herodotus, had predicted that an eclipse would occur no later than the year in which it actually occurred. Herodotus goes on to say that Thales had helped Croesus to divert the river Halys so that his army could cross it.

These anecdotes show that Thales had both scientific and practical interests. His prediction of a solar eclipse, which, according to the astronomers, occurred in 585 BCE, seems quite remarkable, even if, as Herodotus says, he gave only a period of several years in which the eclipse was to occur. Although solar eclipses occur regularly, they are visible only over small portions of the earth, so that their regularity is difficult to discover and verify. Lunar eclipses exhibit the same period as solar eclipses and are easier to observe. Eclipses recur in cycles of about 19 solar years, a period that seems to have been known to many ancient peoples. Among the cuneiform tablets from Mesopotamia, there are many that discuss astronomy, and Ptolemy uses Mesopotamian observations in his system of astronomy. Thales could have acquired this knowledge, along with certain simple facts about geometry, such as the fact that the base angles of an isosceles triangle are equal. Bychkov (2001) argues that the recognition that the base angles of an isosceles triangle are equal probably did come

³The documents from which all this semi-legendary history is assembled are widely scattered. Plutarch, in his *Discourses on the Seven Sages*, Stephanus page 147, section A, said that Thales drove a stake into the ground and used the proportion between the shadows of the stake and the pyramid to compute the height. Diogenes Laertius, in his *Lives of Eminent Philosophers*, Book 1, Section 27, reported a statement by the philosopher Hieronymus of Rhodes (third century BCE) that Thales waited until the length of his own shadow equaled his own height, then measured the length of the shadow of the Great Pyramid. He also reported (Book 1, section 24) the first-century Roman historian Pamphila as saying that Thales was the first to inscribe a right triangle in a circle. He went on to say that others attribute this construction to Pythagoras. (As a matter of general information, Stephanus pagination refers to a definitive sixteenth-century edition of the works of Plato and Plutarch by Henri Estienne (ca. 1530–1598), whose Latin name was Henricus Stephanus.)

from Egypt. In construction—for example, putting a roof on a house—it is not crucial that the cross section be exactly an isosceles triangle, since it is the horizontal edge of the roof that must fit precisely, not the two slanting edges. But when a symmetric square pyramid is built, errors in the base angles of the faces would make it impossible for the faces to fit together tightly along the four oblique edges. Therefore, he believes, Thales must have derived this theorem from his travels in Egypt.

The history of Greek geometry up to the time of Euclid (300 BCE) was written by Eudemus, a pupil of Aristotle. This history was lost, but it is believed to be the basis of the first paragraph of a survey given by Proclus in the fifth century CE in the course of his commentary on the first book of Euclid. In this passage, Proclus mentions 25 men who were considered to have made significant contributions to mathematics. Of these 25, five are well known as philosophers (Thales, Pythagoras, Anaxagoras, Plato, and Aristotle); three are famous primarily as mathematicians and astronomers (Euclid, Eratosthenes, and Archimedes). The other 17 have enjoyed much less posthumous fame. Some of them are so obscure that no mention of them can be found anywhere except in Proclus' summary. Some others (Theodorus, Archytas, Menaechmus, Theaetetus, and Eudoxus) are mentioned by other commentators or by Plato. The 13 just named are the main figures we shall use to sketch the history of Greek geometry. It is clear from what Proclus writes that something important happened to mathematics during the century of Plato and Aristotle, and the result was a unique book, Euclid's *Elements*.

Missing from the survey of Proclus is any reference to Mesopotamian influence on Greek geometry. This influence is shown clearly in Greek astronomy, in the use of the sexagesimal system of measuring angles and in Ptolemy's explicit use of Mesopotamian astronomical observations. It *may* also appear in Book 2 of Euclid's *Elements*, which contains geometric constructions equivalent to certain algebraic relations that are frequently encountered in the cuneiform tablets. This relation, however, is controversial. Leaving aside the question of Mesopotamian influence, we do see a recognition of the Greek debt to Egypt. (Recall Herodotus' conjecture on the origin of Greek geometry from Chapter 7. Euclid actually lived in Egypt, and the other two of the "big three" Greek geometers, Archimedes and Apollonius, both studied there, in the Hellenistic city of Alexandria at the mouth of the Nile.)

8.2.1. Pythagoras

By the time of Pythagoras, geometric lore had expanded beyond the propositions ascribed to Thales, and later commentators constructed an elaborate scenario of a Pythagorean school that devoted itself to the contemplation of geometry and number theory. Although there certainly was a school of Pythagoreans, and Pythagoras was a real person, recent scholarship has cast doubt on its connection with mathematics. In particular, a close analysis of the best-attested citations of Pythagorean doctrine by Burkert (1962) yields a picture of a school preoccupied with mysticism and personal discipline, but not necessarily mathematics. Nevertheless, there remain numerous attributions of mathematical results to the Pythagoreans in the works of the later commentators. Is it possible that Burkert's observation is actually a matter of selection bias on the part of the people who made the quotations? Perhaps these quotations from Pythagoras were chosen by people for whom mathematics was not a priority. It is not certain that all the other attributions of mathematical results to the Pythagoreans are spurious. Even if we grant the possibility that the geometers who assembled the systematic knowledge ascribed to the Pythagoreans actually worked in Athens,

perhaps in Plato's Academy during the early fourth century BCE, after the demise of the original group of Pythagoreans and that these attributions are legends not corresponding to fact, the mere existence of so many attributions makes the name *Pythagorean mathematics* useful as a general description of this pre-Euclidean mathematics.

The philosopher Pythagoras was born on the island of Samos, another of the Greek colonies in Ionia, about half a century after Thales. No books of Pythagoras survive, but many later writers mention him, including Aristotle. Diogenes Laertius devotes a full chapter to the life of Pythagoras. He acquired even more legends than Thales. According to Diogenes Laertius, who cites the grammarian Apollodorus of Athens (ca. 180–ca. 120), Pythagoras sacrificed 100 oxen when he discovered the theorem that now bears his name. If the stories about Pythagoras can be believed, he, like Thales, traveled widely, to Egypt and Mesopotamia. He gathered about him a large school of followers, who observed a mystical discipline and devoted themselves to contemplation. They lived in at least two places in Italy, first at Croton, then, after being driven out,⁴ at Metapontion, where he died in the early fifth century BCE.

According to Book I, Chapter 9 of *Attic Nights*, by the Roman writer Aulus Gellius (ca. 130–180), the Pythagoreans first looked over potential recruits for physical signs of being educable. Those they accepted were first classified as *akoustikoi* (auditors) and were compelled to listen without speaking. After making sufficient progress, they were promoted to *mathēmatikoi* (learners).⁵ Finally, after passing through that state they became *physikoi* (natural philosophers). In his book *On the Pythagorean Life*, Iamblichus uses these terms to denote the successors of Pythagoras, who split into two groups, the *akoustikoi* and the *mathēmatikoi*. According to Iamblichus, the *mathēmatikoi* recognized the *akoustikoi* as genuine Pythagoreans, but the sentiment was not reciprocated. The *akoustikoi* kept the pure Pythagorean doctrine and regarded the *mathēmatikoi* as followers of a disgraced former Pythagorean named Hippasus. This part of the legend probably arose from a passage in Chapter 18, Section 88 of *On the Pythagorean Life*, in which Iamblichus says that Hippasus perished at sea, a punishment for his impiety because he published “the sphere of the 12 pentagons” (probably the radius of the sphere circumscribed about a dodecahedron), taking credit as if he had discovered it, when actually everything was a discovery of That Man (Pythagoras, who was too august a personage to be called by name). Apparently, new knowledge was to be kept in-house as a secret of the initiated and attributed in a mystical sense to Pythagoras.

Diogenes Laertius quotes the philosopher Alexander Polyhistor (ca. 105–35 BCE) as saying that the Pythagoreans generated the world from *monads* (units). By adding a single monad to itself, they generated the natural numbers. By allowing the monad to move, they generated a line, then by further motion the line generated plane figures (polygons), and the plane figures then moved to generate solids (polyhedra). From the regular polyhedra they generated the four elements of earth, air, fire, and water.

From all these sources, one can see how a consistent picture arose of Pythagoreans devoted to understanding the universe mathematically. Despite this plethora of independent sources from ancient times, this picture is not quite consistent with other documents from

⁴Like modern cults, the Pythagoreans seem to have attracted young people, to the despair of their parents. Accepting new members from among the local youth probably aroused the wrath of the citizenry.

⁵Gellius remarks at this point that the word *mathēmatikoi* was being inappropriately used in popular speech to denote a “Chaldean” (astrologer, from a common association with the Chaldean civilization).

the schools of Plato and Aristotle, which indicate that the original Pythagorean group disappeared not long after the death of Pythagoras himself. For that reason, we shall use the word *Pythagorean* sparingly. But we shall use it, since so many ancient authors accepted this view of the history of the subject and wrote as if it were true.

From Proclus and other later authors we have a picture of a sophisticated Pythagorean geometry, entwined with mysticism. For example, Proclus reports that the Pythagoreans regarded the right angle as ethically and aesthetically superior to acute and obtuse angles, since it was “upright, uninclined to evil, and inflexible.” Right angles, he says, were referred to the “immaculate essences,” while the obtuse and acute angles were assigned to divinities responsible for changes in things. The Pythagoreans had a bias in favor of the eternal over the changeable, and they placed the right angle among the eternal things, since unlike acute and obtuse angles, it cannot change without losing its character. In taking this view, Proclus is being a strict Platonist, because Plato’s ideal forms were defined precisely by their absoluteness; they were incapable of undergoing any change without losing their identity.

8.2.2. Mathematical Aspects of Plato’s Philosophy

Plato was interested in mathematics for both philosophical and political reasons. He wanted to solve the crucial problem of governing a state and keeping it stable. To that end, he knew that those with political power needed to understand natural science, and he hoped to provide a “theory of everything,” based on fundamental concepts perceived by the mind, that could be understood by every educated person. Plato is famous for his theory of ideas, which had both metaphysical and epistemological aspects. The metaphysical aspect was a response to two of his predecessors, Heraclitus of Ephesus (ca. 535–475 BCE), who asserted that everything is in constant flux, and Parmenides (born around 515 BCE), who asserted that knowledge is possible only in regard to things that do not change. One can see the obvious implication: Everything changes (Heraclitus). Knowledge is possible only about things that do not change (Parmenides). *Therefore. . .* . To avoid the implication that no knowledge is possible, Plato restricted the meaning of Heraclitus’ “everything” to objects of sense and invented eternal, unchanging forms (ideas) that could be objects of knowledge.

The epistemological aspect of Plato’s philosophy involves universal propositions, statements such as “Lions are carnivorous” (our example, not Plato’s), meaning “*All* lions are carnivorous.” This sentence is grammatically (syntactically) inconsistent with its meaning (semantics). The grammatical subject is the set of all lions, while the assertion is not about this set but about each of its individual members. It asserts that each of them is a carnivore, and therein lies the epistemological problem. *What is the real semantic subject of this sentence, as opposed to the syntactical subject, which is the phrase All lions?* It is not any particular lion. Plato tried to solve this problem by inventing the form or idea of a lion. He would have said that the sentence really asserts a relation perceived in the mind between the form of a lion and the form of a carnivore. Mathematics, because it dealt with objects and relations perceived by the mind, appeared to Plato to be the bridge between the world of sense and the world of forms. Nevertheless, mathematical objects were not the same thing as the forms. Each form, Plato claimed, was unique. Otherwise, the interpretation of universal propositions by use of forms would be ambiguous. But mathematical objects such as lines are not unique. There must be at least three lines, for example, in order for a triangle to exist. Hence, as a sort of hybrid of sense experience and pure mental creation, mathematical objects offered a way for the human soul to ascend to the height of understanding, by perceiving the forms themselves. Incorporating mathematics into education so as to

realize this program was Plato's goal, and his pupils studied mathematics in order to achieve it. Although the philosophical goal was not reached, the effort expended on mathematics was not wasted; certain geometric problems were solved by people associated with Plato, providing the foundation of Euclid's famous work, known as the *Elements*.

A little over half a century after Plato's death, Euclid wrote his famous treatise, the *Elements*, which is quite free of all the metaphysical distractions that had preoccupied Plato. Later, neo-Platonic philosophers such as Proclus attempted to reintroduce philosophical ideas into their commentary on Euclid's work. Neugebauer (1975, p. 572) described the philosophical aspects of Proclus' introduction as "gibberish," and expressed relief that scientific methodology survived despite the prevalent dogmatic philosophy.

According to Diels (1951, 44A5), Plato met the Pythagorean Philolaus in Sicily in 390. In any case, Plato must certainly have known the work of Philolaus, since in the *Phaedo*, Socrates says that both Cebes and Simmias are familiar with the work of Philolaus and implies that he himself knows of it at second hand. It seems likely, then, that Plato's interest in mathematics began some time after the death of Socrates and continued for the rest of his life, that mathematics played an important role in the curriculum of his Academy and the research conducted there, and that Plato himself played a role in directing that research. We do not, however, have any theorems that can with confidence be attributed to Plato himself. Lasserre (1964, p. 17) believed that the most important mathematical work at the Academy was done between 375 and 350 BCE.

In Book VII of Plato's *Republic*, Socrates explained that arithmetic was needed both to serve the eye of the soul and as a practical instrument in planning civic projects and military campaigns:

The kind of knowledge we are seeking seems to be as follows. It is necessary for a military officer to learn (*matheîn*) these things for the purpose of proper troop deployment, and the philosopher must have risen above change, in order to grasp the essence of things, or else never become skilled in calculation (*logistikós*).

Later in the same book, Plato, through Socrates, complains of the lack of a government subsidy for geometry. In his day, solid geometry was underdeveloped in comparison with plane geometry, and Socrates gave what he thought were the reasons for its backwardness:

First, no government holds [the unsolved problems in solid geometry] in honor; and they are researched in a desultory way, being difficult. Second, those who are doing the research need a mentor, without which they will never discover anything. But in the first place, to become a mentor is difficult; and in the second place, after one became a mentor, as things are just now, the arrogant people doing this research would never listen to him. But if the entire state were to act in concert in conducting this research with respect, the researchers would pay heed, and by their combined intensive work the answers would become clear.

Plato himself, although he had practical objects in mind, connected with the best possible government, was also an intellectual for whom the "eye of the soul" was sufficient justification for intellectual activity. He seems to have had a rather dim view of purely practical-minded people. In his long dialogue *The Laws*, one of the speakers, an Athenian, rants about the shameful Greek ignorance of incommensurables, surely a topic of limited application in the lives of most people. (Plato would probably say even worse things about the modern world, where almost no one knows what incommensurables are!)

8.3. WORKS AND AUTHORS

Extensive treatises on mathematics written in Greek began appearing early in the Hellenistic era (third century BCE) and continued in a steady stream for hundreds of years. We list here only a few of the most outstanding authors.

8.3.1. Euclid

This author lived and worked in Alexandria, having been invited by Ptolemy Soter (Ptolemy I) shortly after the city was founded. At the 2006 winter meeting of the Canadian Mathematical Society, historian Alexander Jones argued that Euclid probably flourished around the middle of the third century and was a contemporary of Archimedes. Essentially nothing is known of his life beyond the fact that he worked in Alexandria, but his famous treatise on the basics of geometry (the *Elements*) has become a classic known all over the world. Several of his minor works—the *Optics*, the *Data*, and the *Phaenomena*—also have been preserved. Euclid did not provide any preface to tell us why he wrote his treatise. We do, however, know enough of the Platonic philosophy to understand why he developed geometry and number theory to the extent that he did, and it is safe to conclude that this kind of work was considered valuable because it appealed to the intellect of those who could understand it.

8.3.2. Archimedes

Much more is known of Archimedes (ca. 287–212 BCE). About 10 of his works have been preserved, including the prefaces that he wrote in the form of “cover letters” to the people who received the works. Here is one such letter, which accompanied a report of what may well be regarded as his most profound achievement—proving that the surface of a sphere is four times as large as its equatorial disk.

On a former occasion I sent you the investigations which I had up to that time completed, including the proofs, showing that any segment bounded by a straight line and a section of a right-angled cone [parabola] is four-thirds of the triangle which has the same base with the segment and equal height. Since then certain theorems not hitherto demonstrated have occurred to me, and I have worked out the proofs of them. They are these: first, that the surface of any sphere is four times its greatest circle. . . . For, though these properties also were naturally inherent in the figures all along, yet they were in fact unknown to all the many able geometers who lived before Eudoxus, and had not been observed by anyone. Now, however, it will be open to those who possess the requisite ability to examine these discoveries of mine. [Heath, 1897, Dover edition, pp. 1–2]

As this letter shows, mathematics was a “going concern” by Archimedes’ time, and a community of mathematicians existed. Archimedes is known to have studied in Alexandria. He perished when his native city of Syracuse was taken by the Romans during the Second Punic War. Some of Archimedes’ letters, like the one quoted above, give us a glimpse of mathematical life during his time. Despite being widely separated, the mathematicians of the time sent one another challenges and communicated their achievements.

8.3.3. Apollonius

Apollonius, about one generation younger than Archimedes, was a native of what is now Turkey. He studied in Alexandria after the time of Euclid and is also said to have taught there. He eventually settled in Pergamum (now Bergama in Turkey). He is the author of eight books on conic sections, four of which survive in Greek and three others in an Arabic translation. We know that there were originally eight books because commentators, especially Pappus, described the work and reported the number of propositions in each book.

In his prefaces, Apollonius implies that geometry was simply part of what an educated person would know, and he also implies that such people were as fascinated with it in his time as they are today about the latest scientific achievements. Among other things, he said the following.

During the time I spent with you at Pergamum I observed your eagerness to become acquainted with my work in conics. [Book I]

I undertook the investigation of this subject at the request of Naucrates the geometer, at the time when he came to Alexandria and stayed with me, and, when I had worked it out in eight books, I gave them to him at once, too hurriedly, because he was on the point of sailing; they had therefore not been thoroughly revised, indeed I had put down everything just as it occurred to me, postponing revision until the end. [Book II]

8.3.4. Zenodorus

Zenodorus (second century BCE) represents a new departure in the Euclidean tradition. Instead of proving direct proportions, as earlier mathematicians had done, he worked with inequalities and showed, as well as could be done given the tools available to him, that a regular polygon encloses a larger area than any other polygon of the same perimeter and the same number of sides, that the more sides a regular polygon of a given perimeter has, the greater the area it encloses, and that a circle encloses a larger area than any polygon whose perimeter equals the circumference of the circle. He also established similar theorems for polyhedra and spheres. These *isoperimetric problems* are not found in Euclid or Apollonius, and Archimedes only hints at them when he points out the need to assume that a convex curve enclosing another convex curve must be longer than the one it encloses. These results of Zenodorus are known because Theon of Alexandria quoted them in his commentary on Ptolemy's *Sýntaxis*. Pappus borrowed freely from Zenodorus in his own work on such problems.

8.3.5. Heron

This mathematician and engineer (first century CE) is also known as Hero (just as Plato is actually known in Greek as Platon). The name Heron was very common in his world, and it is difficult to be sure that a person by that name is any particular Heron one might have in mind. The one we shall be discussing is famous for having invented a steam engine of sorts, but we shall be interested only in the way that he represents the return of metric concepts to geometry, using numbers to describe the lengths of the sides of a triangle, for example, and giving a method of computing the area of a triangle knowing the lengths of its sides.

8.3.6. Ptolemy

Claudius Ptolemy was primarily an astronomer and physicist, although these subjects were hardly distinct from mathematics in his time. He lived in Alexandria during the second century, as is known from the astronomical observations that he made between 127 and 141 CE. He created an intricate and workable earth-centered mathematical system of explaining the motion of the planets and systematized it in a treatise known as the *Sýntaxis* (treatise, literally *arrangement*), which consisted of 13 books. Ptolemy's *Sýntaxis* became a classic reference and was used for well over a thousand years as the definitive work on mathematical astronomy. It became known as the "greatest" work (*megístē* in Greek) on astronomy and, when translated into Arabic, became *al-megista* or the *Almagest*, as we know it today.

8.3.7. Diophantus

Little is known about this author of a remarkable treatise on what we now call algebra and number theory. He probably lived in the third century CE, although some experts believe he lived earlier than that. His treatise is of no practical value in science or commerce, but its problems inspired number theorists during the seventeenth century and led to the long-standing conjecture known as Fermat's last theorem. The 1968 discovery of what may be four books from this treatise that were long considered lost was the subject of a debate among the experts, some of whom believed the books might be commentaries, perhaps written by the late fourth-century commentator Hypatia. If so, they would be the only work by Hypatia still in existence.

8.3.8. Pappus

Pappus, who is known to have observed a solar eclipse in Alexandria in 320 CE, was the most original and creative of the later commentators on Greek geometry and arithmetic. His *Synagōgē* (*Collection*) consists of eight books of insightful theorems on arithmetic and geometry, as well as commentary on the works of other authors. In some cases where works of Euclid, Apollonius, and others have been lost, this commentary tells something about these works. Pappus usually writes as if the reader will have a natural interest in his subject matter, but occasionally he gives in addition a practical justification for his study, as in Book 8:

The science of mechanics, my dear Hermodorus, has many important uses in practical life, and is held by philosophers to be worthy of the highest esteem, and is zealously studied by mathematicians, because it takes almost first place in dealing with the nature of the material elements of the universe. [Thomas, 1941, p. 615]

8.3.9. Theon and Hypatia

The later commentators Theon of Alexandria (late fourth century) and his daughter Hypatia (ca. 370–415) also produced respectable work, including a standard edition of Euclid's *Elements*. Several of Theon's commentaries still exist, but nothing authored by Hypatia has been preserved, unless the books of Diophantus mentioned above were written by her. Very

little of value can be found in Greek mathematics after the fourth century. As Gow (1884, p. 308) says:

The *Collection* of Pappus is not cited by any of his successors, and none of them attempted to make the slightest use of the proofs and *aperçus* in which the book abounds. . . His work is only the last convulsive effort of Greek geometry which was now nearly dead and was never effectually revived.

QUESTIONS

Historical Questions

- 8.1. Describe in general terms the periods of development, flourishing, and decline in ancient Greek mathematics, naming the primary authors and their works.
- 8.2. Who are the commentators who provide the context of the major works of ancient Greek mathematics?
- 8.3. In what way does ancient Greek mathematics differ from the mathematics of Mesopotamia and Egypt?

Questions for Reflection

- 8.4. What advantages can you see in an axiomatic development of mathematics starting from definitions and assumptions? Are there disadvantages?
- 8.5. Given that we have no documents from the time of Greek mathematics—the earliest manuscripts we have are medieval—how can we be sure that the texts we have are actually what the authors wrote? Were the copyists who wrote the early medieval manuscripts simply concerned with reproducing the text faithfully, or is it possible that they tried to improve it by revisions they thought of themselves? How could we know if they did?
- 8.6. Plato thought that mathematics was a sort of entranceway into the ideal world of his forms; and he also thought that the physical world, though corrupt, could be understood by relations grasped by the mind rather than the senses. To what extent is this view plausible? Does modern theoretical physics presume something similar?

Greek Number Theory

Greek number theory is of interest both intrinsically, because some of the natural questions that it raised have not been answered even in the present time, and because it was the soil in which algebraic symbolism first sprouted, a brilliant innovation during a time otherwise marked by intellectual decline. The theory itself has two areas of interest that do not interact during the period of Greek intellectual dominance. One is the Pythagorean topic of the arithmetic properties of figurate numbers (triangular numbers, square numbers, pentagonal numbers, and so on). This area has declined greatly in importance, along with Pythagoreanism and neo-Platonism, although it is not quite extinct even today. The recent solution of the problem of Fermat's last theorem is a good specimen of the modern development of this theory. The other area, which provides the theoretical foundation for much of classical and modern number theory, is the theory of divisibility of integers. This area also has one rather Pythagorean connection, namely the topic of perfect numbers. We shall look at just three of the classical Greek writers on number theory:

1. Euclid, whose *Elements* contain three books (Books 7–9) devoted to the divisibility properties of integers.
2. Nicomachus of Gerasa, a neo-Pythagorean philosopher who lived about 100 CE. His *Introduction to Arithmetic* gives a detailed development of both figurate numbers and the divisibility theory.
3. Diophantus, who lived sometime between the second and fourth centuries CE. He is sometimes justly called the “father of algebra,” because in the course of his study of the arithmetic properties of square numbers, he introduced symbolic notation for an unspecified or unknown number and a way of writing operations on unknown numbers.

As it happens that the treatise of Nicomachus preserves more of what was traditionally called Pythagorean lore than the earlier work of Euclid, we shall discuss Nicomachus first. Before we can do that, however, we need to digress to explain a mathematical technique that is fundamental to the understanding of a great deal in the history of mathematics.

9.1. THE EUCLIDEAN ALGORITHM

The Greeks learned early on how to find the greatest common divisor of two numbers. A very efficient procedure for doing so is described in Chapter 13 of Book 1 of Nicomachus' *Arithmetica* and in Proposition 2 of Book 7 of Euclid's *Elements*. This procedure is now known as the *Euclidean algorithm*, Nicomachus applies it only to integers, any two of which naturally have 1 as a common divisor. Euclid, on the other hand, does not confine it to integers, but states the procedure for "magnitudes," which may lack a common measure. It is significant that when it is applied to continuous magnitudes, the procedure terminates if and only if there is a common measure. Euclid makes use of that fact in discussing incommensurables, which are pairs of magnitudes having no common measure. The algorithm was certainly invented long before the time of Euclid, however. Zverkina (2000) believes that this procedure could not have arisen intuitively, but must have come about as the result of solving specific problems, most likely the problem of reducing ratios by canceling a common divisor. What follows is a description of the general procedure.

For definiteness, we shall imagine that the two quantities whose greatest common measure is to be found are two lengths, say a and b . Suppose that a is longer than b . (If the two are equal, their common value is also their greatest common divisor.) The general procedure is to keep subtracting the smaller quantity from the larger until the remainder is equal to the smaller quantity or smaller than it. It is not difficult to show that the smaller quantity and the remainder have the same common measures as the smaller quantity and the larger. Hence one can start over with the smaller quantity and the remainder, which is no more than half of the larger quantity. Either this process terminates with an equal pair, or it continues and the pairs become arbitrarily small.

An example using integers will make the procedure clear. Let us find the greatest common measure (divisor) of 26173996849 and 180569389. A common measure does exist: the integer 1. Since the repeated subtraction process amounts to division with remainder, we do it this way: $26173996849 \div 180569389$ is 144 with a remainder of 172004833. We then divide the smaller quantity (the previous divisor) 180569389 by the remainder 172004833, getting a quotient of 1 and a remainder of 8564556. Next we divide the previous remainder 172004833 by the new remainder 8564556, getting a quotient of 20 and a remainder of 713713. We then divide 8564556 by 713713 and get a quotient of 12 with no remainder, so that the greatest common divisor is 713713.

This computation can be arranged as follows, with the successive divisions performed from right to left. The greatest common measure appears at the extreme left:

$$\begin{array}{r}
 26173996849 \\
 \underline{26001992016} \\
 8564556 \\
 \underline{8564556} \\
 0
 \end{array}$$

One can see that this procedure must produce a greatest common measure if one exists, since the first remainder is at most half of the larger of the two original quantities (since it is smaller than the smaller of the two original quantities and not larger than their difference). Similarly, since the smaller quantity becomes the dividend in the second application, the second remainder will be at most half of it. Thus the first two remainders are at most half the size of the original quantities. Yet they are both larger than any common measure the

two original quantities had. It follows that if there is any common measure, one of these remainders will ultimately become zero, since repeated halving would otherwise eventually make it smaller than that common measure. The last nonzero remainder is thus the greatest common measure of the two quantities.

9.2. THE ARITHMETICA OF NICOMACHUS

In his first book, Nicomachus makes the elementary distinction between odd and even numbers. Having made this distinction, he proceeds to refine it, distinguishing between even numbers divisible by 4 (evenly even) and those that are not (doubles of odd numbers). He goes on to classify odd numbers in a similar way, thereby coming to the concept of prime and composite numbers. Nicomachus also introduces what we now call pairs of *relatively prime numbers*. These are pairs of numbers that have no common prime divisor and hence no common divisor except 1. Relational properties were difficult for Greek philosophers, and Nicomachus expresses the concept of relatively prime numbers in a confused manner, referring to three species of odd numbers: the prime and incomposite, the secondary and composite, and “the variety which, in itself, is secondary and composite, but relatively is prime and incomposite.” This way of writing seems to imply that there are three kinds of integers, prime and incomposite, secondary and composite, and a third kind midway between the other two. It also seems to imply that one can look at an individual integer and classify it into exactly one of these three classes. Such is not the case, however. The property of primeness is a property of a number alone. The property of being relatively prime is a property of a pair of numbers. On the other hand, the property of being relatively prime to a given number is a property of a number alone. Nicomachus explains the property in a rather wordy fashion in Chapter 13 of Book 1, where he gives a method of identifying prime numbers that has become famous as the *sieve of Eratosthenes*.

Nicomachus attributes this method to Eratosthenes (276–174 BCE, best known for this work on prime numbers and for having estimated the size of the earth). To use it, start with a list of all the odd numbers from 3 on, that is,

3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37,

From this list, remove the multiples of 3, starting with $3 \cdot 3$, that is, remove 9, 15, 21, 27, 33, The reduced list is then

3, 5, 7, 11, 13, 17, 19, 23, 25, 29, 31, 35, 37, 41, 43, 47, 49,

From this new list, remove all multiples of 5, starting with $5 \cdot 5$. The first nonprime in the resulting list will $49 = 7 \cdot 7$, and so you remove all multiples of 7 from that list. In this way, you can generate in short order a complete list of primes up to the square of the first prime whose multiples were not removed. Thus, after removing the multiples of 7, we have the list

3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61

The first nonprime in this list would be $11 \cdot 11 = 121$.

9.2.1. Factors vs. Parts. Perfect Numbers

Nicomachus' point of view on this sieve was different from ours. Where we think of the *factors* of, say 60, as being 2, 2, 3, and 5, Nicomachus thought of the quotients on division by these factors and products of these factors as the *parts* of a number. Thus, in his language, 60 has the parts 30 (half of 60), 20 (one-third of 60), 15 (one-fourth of 60), 12 (one-fifth of 60), 10 (one-sixth of 60), 6 (one-tenth of 60), 5 (one-twelfth of 60), 4 (one-fifteenth of 60), 3 (one-twentieth of 60), 2 (one-thirtieth of 60), and 1 (one-sixtieth of 60). If these parts are added, the sum is 108, much larger than 60. Nicomachus called such a number *superabundant* and compared it to an animal having too many limbs. On the other hand, 14 is larger than the sum of its parts. Indeed, it has only the parts 7, 2, and 1, which total 10. Nicomachus called 14 a *deficient number* and compared it to an animal with missing limbs like the one-eyed Cyclops of the *Odyssey*. A number that is exactly equal to the sum of its parts, such as $6 = 1 + 2 + 3$, he called a *perfect number*. He gave a method of finding perfect numbers, which remains to this day the only way known to generate such numbers, although it has not been proved that there are no other such numbers. This procedure is also stated by Euclid as Proposition 36 of Book 9 of the *Elements*: *If the sum of the numbers 1, 2, 4, . . . , 2^{n-1} is prime, then this sum multiplied by the last term will be perfect.* To see the recipe at work, start with 1, then double and add: $1 + 2 = 3$. Since 3 is prime, multiply it by the last term, that is, 2. The result is 6, a perfect number. Continuing, $1 + 2 + 4 = 7$, which is prime. Multiplying 7 by 4 yields 28, the next perfect number. Then, $1 + 2 + 4 + 8 + 16 = 31$, which is prime. Hence $31 \cdot 16 = 496$ is a perfect number. The next such number is $8128 = 64(1 + 2 + 4 + 8 + 16 + 32 + 64)$. In this way, Nicomachus was able to generate the first four perfect numbers. He seems to hint at a conjecture, but draws back from stating it explicitly:

When these have been discovered, 6 among the units and 28 in the tens, you must do the same to fashion the next. . . the result is 496, in the hundreds; and then comes 8,128 in the thousands, and so on, as far as it is convenient for one to follow [D'ooge, 1926, p. 211].¹

This quotation seems to imply that Nicomachus expected to find one perfect number N_k having k decimal digits. Actually, the fifth perfect number is 33,550,336, so we have jumped from four digits to eight here. The sixth is 8,589,869,056 (10 digits) and the seventh is 137,438,691,328 (12 digits), so that there is no regularity about the distribution of perfect numbers. Thus, Nicomachus was wise to refrain from making conjectures too explicitly. According to Dickson (1919, p. 8), later mathematicians, including the great sixteenth-century algebraist Girolamo Cardano, were less restrained, and this incorrect conjecture has been stated more than once.

For a topic that is devoid of applications, perfect numbers have attracted a great deal of attention from mathematicians. Dickson (1919) lists well over 100 mathematical papers devoted to this topic over the past few centuries. From the point of view of pure number theory, the main questions about them are the following: (1) Is there an odd perfect number?² (2) Are all even perfect numbers given by the procedure described by

¹D'ooge illustrates the procedure in a footnote, but states erroneously that 8191 is not a prime.

²The answer is unknown at present.

Nicomachus?³ (3) Which numbers of the form $2^n - 1$ are prime? These are called *Mersenne primes*, after Marin Mersenne (1588–1648), who, according to Dickson (1919, pp. 12–13), first noted their importance, precisely in connection with perfect numbers. Obviously, n must itself be prime if $2^n - 1$ is to be prime, but this condition is not sufficient, since $2^{11} - 1 = 23 \cdot 89$. The set of known prime numbers is surprisingly small, considering that there are infinitely many to choose from, and the new ones being found tend to be Mersenne primes, mostly because that is where people are looking for them. The largest currently known prime, discovered on August 23, 2008, is $2^{43112609} - 1$, only the forty-fifth Mersenne prime known at the time. Since then, two more have been discovered, both smaller than this one however.⁴ It was found by the GIMPS (Great Internet Mersenne Prime Search) project, which links hundreds of thousands of computers via the Internet and runs prime-searching software in the background of each while their owners are busy with their own work. This prime has 12,978,189 decimal digits. Since these primes are not being discovered in ascending order, it is not accurate to call the largest currently known one the 47th Mersenne prime. Exhaustive checking by the GIMPS network since the fortieth Mersenne prime, $2^{20996011} - 1$, was discovered on November 17, 2003 (it has 6,320,430 decimal digits) has established that there are no others smaller than that one. Thus we know the first 40 Mersenne primes and seven others as well. In contrast, the largest non-Mersenne prime known as of late 2011 was $19249 \cdot 2^{13018586} + 1$, discovered in May 2007; it has 3,918,990 decimal digits and hence is tiny compared with the largest known Mersenne primes. (This information comes from the website <http://primes.utm.edu>.)

9.2.2. Figurate Numbers

Beginning in Chapter 6 of Book 2, Nicomachus studies figurate numbers: polygonal numbers through heptagonal numbers, and then polyhedral numbers. These numbers are connected with geometry, the number 1 being replaced by a geometric point. To motivate this discussion, Nicomachus speculated that the simplest way to denote any integer would be repeating a symbol for 1 an appropriate number of times. Thus, he said, the number 5 could be denoted $\alpha \alpha \alpha \alpha \alpha$. This train of thought, if followed consistently, would lead back to a notation even more primitive than the hieroglyphic notation for numbers, since it would use only the symbol for units and discard the symbols for higher powers of 10. The Egyptians had gone beyond this principle in their hieratic notation, and the standard Greek notation was essentially a translation of the hieratic into the Greek alphabet. You can easily see where this speculation leads. The outcome is shown in Fig. 9.1, which illustrates triangular, square, pentagonal, and hexagonal numbers using dots instead of the letter α . Observe that the figures are *not* associated with regular polygons except in the case of triangles and squares. The geometry alone makes it clear that a square number is the sum of the corresponding triangular number and its predecessor. Similarly, a pentagonal number is the sum of the corresponding square number and the preceding triangular number, a hexagonal number is the sum of the corresponding pentagonal number and the preceding triangular number, and so forth. This is the point at which modern mathematics parts company with Nicomachus,

³The answer is yes. The result is amazingly easy to prove, but no one seems to have noticed it until a posthumous paper of Leonhard Euler gave a proof. Victor-Amédée Lebesgue (1791–1875) published a short proof in 1844.

⁴The reader will correctly infer from previous footnotes that exactly 47 perfect numbers are now known.

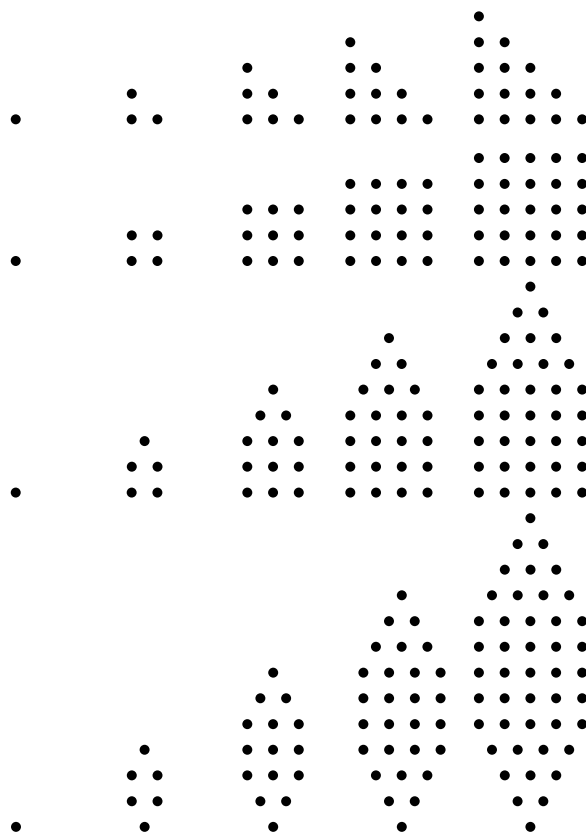


Figure 9.1. Figurate numbers. Top row: triangular numbers $T_n = n(n + 1)/2$. Second row: square numbers $S_n = n^2$. Third row: pentagonal numbers $P_n = n(3n - 1)/2$. Bottom row: hexagonal numbers $H_n = n(2n - 1)$.

Proclus, and other philosophers who push analogies further than the facts will allow. As Nicomachus states at the beginning of Chapter 7:

The point, then, is the beginning of dimension, but not itself a dimension, and likewise the beginning of a line, but not itself a line; the line is the beginning of surface, but not surface; and the beginning of the two-dimensional, but not itself extended in two dimensions. . . Exactly the same in numbers, unity is the beginning of all number that advances unit by unit in one direction; linear number is the beginning of plane number, which spreads out like a plane in one more dimension. [D'ooze, 1926, p. 239]

This mystical mathematics was transmitted to Medieval Europe by Boethius. It is the same kind of analogical thinking found in Plato's *Timaeus*, where it is imagined that atoms of fire are tetrahedra, atoms of earth are cubes, and so forth. Since the Middle Ages, this topic has been of less interest to mathematicians. The phrase *of less interest*—rather than *of no interest*—is used advisedly here: There are a few theorems about figurate numbers in modern number theory, and they have some connections with analysis as well.

For example, a formula of Euler asserts that

$$\prod_{k=1}^{\infty} (1 - x^k) = \sum_{n=-\infty}^{\infty} (-1)^n x^{n(3n-1)/2}.$$

Here the exponents on the right-hand side range over the pentagonal numbers for n positive. By defining the n th pentagonal number for negative n to be $n(3n - 1)/2$, we gain an interesting formula that can be stated in terms of figurate numbers. Carl Gustav Jacobi (1804–1851) was pleased to offer a proof of this theorem as evidence of the usefulness of elliptic function theory. Even today, these numbers crop up in occasional articles in graph theory and elsewhere.

9.3. EUCLID'S NUMBER THEORY

Euclid devotes most of his three books on number theory to divisibility theory, spending most of the time on proportions among integers and on prime and composite numbers, with fewer results on figurate numbers. Only at the end of Book 9 does he prove a theorem of a different sort, giving the method of constructing perfect numbers described above. It is interesting that, except for squares and cubes, Euclid does not mention figurate numbers. Although the Pythagorean and Platonic roots of Euclid's treatise are obvious, Euclid appears to the modern eye to be much more a mathematician than Pythagoras or Plato, not at all inclined to flights of fanciful speculation on the nature of the universe. In fact, he never mentions the universe at all and suggests no practical applications of the theorems in his *Elements*.

Book 7 develops proportion for positive integers as part of a general discussion of ways of reducing a ratio to lowest terms. The notion of relatively prime numbers is introduced, and the elementary theory of divisibility is developed as far as finding least common multiples and greatest common factors. Book 8 resumes the subject of proportion and extends it to squares and cubes of integers, including the interesting theorem that the mean proportional of two square integers is an integer (Proposition 11—for example, $25 : 40 :: 40 : 64$), and between any two cubes there are two such mean proportionals (Proposition 12—for example, $27 : 45 :: 45 : 75 :: 75 : 125$). Book 9 continues this topic; it also contains the famous theorem that there are infinitely many primes (Proposition 20, in the form of the assertion that no given finite collection of primes can contain all of them) and ends by giving the only known method of constructing perfect numbers (Proposition 36), quoted above.

Euclid's number theory does not contain any explicit statement of the *fundamental theorem of arithmetic* (Knorr, 1976). This theorem, which asserts that every positive integer can be written in only one way as a product of prime numbers, can easily be deduced from Book 7, Proposition 24: *If two numbers are relatively prime to a third, their product is also relatively prime to it.*

9.4. THE ARITHMETICA OF DIOPHANTUS

Two works of Diophantus have survived in part, a treatise on polygonal numbers and the work for which he is best known, the *Arithmetica*. Like many other ancient works,

these two works of Diophantus survived because of the efforts of a ninth-century Byzantine mathematician named Leon, who organized a major effort to copy and preserve them. There is little record of the influence the works of Diophantus may have exerted before this time. What is of particular interest to us is his study of the arithmetic properties of squares. He is interested in finding ways to represent a given rational square number as the sum of two other rational square numbers. The techniques he developed to solve that problem resulted in the first use of symbolism to represent the kind of abstract thinking required in algebra.

To judge this work, one should know something of its predecessors and its influence. Unfortunately, information about either of these is difficult to come by. The Greek versions of the treatise, of which there are 28 manuscripts according to Sesiano (1982, p. 14), all date to the thirteenth century. Among the predecessors of Diophantus, we can count Heron of Alexandria and one very obscure Thymaridas, who showed how to solve a particular set of linear equations, known as the *epanthēma* (blossom) of Thymaridas.

Because the work of Diophantus is so different from the style of Euclid and his immediate successors, the origins of his work have been traced to other cultures, notably Egypt and Mesopotamia. The historian of mathematics Paul Tannery (1843–1904) printed an edition of Diophantus' work and included a fragment supposedly written by the eleventh-century writer Michael Psellus (1018–ca. 1078), which stated that “As for this *Egyptian* method, while Diophantus developed it in more detail, . . .” On this basis, Tannery assigned Diophantus to the third century. Neugebauer (1952, p. 80) distinguished two threads in Hellenistic mathematics, one in the logical tradition of Euclid, the other having roots in the Babylonian and Egyptian procedures and says that, “the writings of Heron and Diophantus. . . form part of this oriental tradition which can be followed into the Middle Ages both in the Arabic and in the western world.” Neugebauer saw Diophantus as reflecting an earlier type of mathematics practiced in Greece alongside the Pythagorean mathematics and temporarily eclipsed by the Euclidean school. As he said (1952, p. 142):

It seems to me characteristic, however, that Archytas of Tarentum could make the statement that not geometry but arithmetic alone could provide satisfactory proofs. If this was the opinion of a leading mathematician of the generation just preceding the birth of the axiomatic method, then it is rather obvious that early Greek mathematics cannot have been very different from the Heronic Diophantine type.

9.4.1. Algebraic Symbolism

Diophantus began by introducing a symbol for a constant unit $\overset{\circ}{M}$, from *monás* (*Μονάς*), along with a symbol for an unknown number ζ , conjectured to be an abbreviation of the first two letters of the Greek word for number: *arithmós* (*ἀριθμός*). For the square of an unknown he used Δ^{ν} , the first two letters of *dýnamis* (*Δύναμις*), meaning *power*. For its cube he used K^{ν} , the first two letters of *kýbos* (*Κύβος*), meaning *cube*. He then combined these letters to get fourth ($\Delta^{\nu}\Delta$), fifth (ΔK^{ν}), and sixth ($K^{\nu}K$) powers. For the reciprocals of these powers of the unknown he invented names by adjoining the suffix *-ton* (*-τον*) to the names of the corresponding powers. These various powers of the unknown were called *eída* (*εἶδα*), meaning *species*. Diophantus' system for writing down the equivalent of a polynomial in the unknown consisted of writing down these symbols in order to indicate addition, each term followed by the corresponding number symbol (for which the Greeks used their alphabet). Terms to be added were placed first, separated by a pitchfork (\pitchfork) from those to be subtracted.

Heath conjectured that this pitchfork symbol is a condensation of the letters lambda and iota, the first two letters of a Greek root meaning *less* or *leave*. Thus what we would call the expression $2x^4 - x^3 - 3x^2 + 4x + 2$ would be written $\Delta^{\nu} \Delta \bar{\beta} \zeta \bar{\delta} \overset{\circ}{M} \bar{\beta} \uparrow K^{\nu} \bar{\alpha} \Delta^{\nu} \bar{\gamma}$.

Diophantus' use of symbolism is rather sparing by modern standards. He often uses words where we would use symbolic manipulation. For this reason, his algebra was described by the nineteenth-century German orientalist Nesselmann (1811–1881) as a transitional “syncopated” phase between the earliest “rhetorical” algebra, in which everything is written out in words, and the modern “symbolic” algebra.

9.4.2. Contents of the *Arithmetica*

According to the introduction to the *Arithmetica*, this work consisted originally of 13 books, but until recently only six were known to have survived. It was assumed that these were the first six books, on which Hypatia was known to have written a commentary. However, more books were recently found in an Arabic manuscript that the experts say is a translation made very early—probably in the ninth century. Sesiano (1982) stated that these books are in fact the books numbered 4 to 7, and that the books previously numbered 4 to 6 must come after them.

Diophantus begins with a small number of determinate problems that illustrate how to think algebraically using the symbolic notation discussed above. Indeterminate problems, which are number theory because the solutions are required to be rational numbers (the only kind recognized by Diophantus), begin in Book 2.⁵ A famous example of this type is Problem 8 of Book 2: *Separate a given square number into two squares*. Diophantus illustrates this problem using the number 16 as an example. His method of solving this problem is to express the two numbers in terms of a single unknown ζ in such a way that one of the conditions is satisfied automatically. Thus, letting one of the two squares be ζ^2 , which Diophantus wrote as Δ^{ν} , he noted that the other will automatically be $16 - \zeta^2$. To get a determinate equation for ζ , he assumes that the other number to be squared is 4 less than an unspecified multiple of ζ . The number 4 is chosen because it is the square root of 16. In our terms, it leads to a quadratic equation one of whose roots is zero, so that the other root can be found by solving a linear equation. As we would write it, assuming that $16 - \zeta^2 = (k\zeta - 4)^2$, we find that $(k^2 + 1)\zeta^2 = 8k\zeta$, and—canceling ζ , since Diophantus does not operate with 0—we get $\zeta = 8k/(k^2 + 1)$. This formula generates a whole infinite family of solutions of the equation that we would call $x^2 + y^2 = 16$ via the identity

$$\left(\frac{8k}{k^2 + 1}\right)^2 + \left(\frac{4(k^2 - 1)}{k^2 + 1}\right)^2 = 16.$$

You may be asking why it was necessary to use a square number (16) here. Why not separate any positive rational number, say 5, into a sum of two squares? If you look carefully at the solution, you will see that Diophantus had to make the constant term drop out of the quadratic equation, and that could only be done by introducing the square root of the given number.

⁵Although Diophantus allowed solutions to be what we now call positive rational numbers, the name *Diophantine equation* is now used to refer to an indeterminate equation or system in which the solutions are required to be integers.

Diophantus' procedure is slightly less general than what we have just shown, although his illustrations show that he knows the general procedure and could generate other solutions. In his illustration he assumes that the other square is $(2\zeta - 4)^2$. Since this number must be $16 - \zeta^2$, he finds that $4\zeta^2 - 16\zeta + 16 = 16 - \zeta^2$, so that $\zeta = \frac{16}{5}$. It is clear that this procedure can be applied very generally, with the coefficient 2 replaced by any positive integer, showing an infinite number of ways of dividing a given square into two other squares.

At first sight it appears that number theory really is not involved in this problem, that it is a matter of pure algebra. This topic, however, naturally leads to other questions that definitely do involve number theory, that is, the theory of divisibility of integers. The most obvious one is the problem of finding *all possible* representations of a positive rational number as the sum of the squares of two rational numbers. One could then generalize and ask how many ways a given rational number can be represented as the sum of the cubes or fourth powers, and so forth, of two rational numbers. Those of a more Pythagorean bent might ask how many ways a number can be represented as a sum of triangular, pentagonal, or hexagonal numbers. In fact, many questions like this have been asked. Leonhard Euler (1707–1783) proved that it was impossible for the sum of fewer than three cubes to equal a cube and conjectured that it was impossible for the sum of fewer than n n th powers to equal another n th power. (He was wrong: It is possible for the sum of four fifth powers to equal a fifth power.)

9.4.3. Fermat's Last Theorem

The problem just solved achieved lasting fame when Fermat, who was studying the *Arithmetica*, remarked that the analogous problem for cubes and higher powers had no solutions; that is, one cannot find positive integers x , y , and z satisfying $x^3 + y^3 = z^3$ or $x^4 + y^4 = z^4$, or, in general, $x^n + y^n = z^n$ with $n > 2$. Fermat stated that he had found a proof of this fact, but unfortunately did not have room to write it in the margin of the book. Fermat never published any general proof of this fact, although the special case $n = 4$ is a consequence of a method of proof developed by Fermat, known as the method of infinite descent. The problem became generally known after 1670, when Fermat's son published an edition of Diophantus' work along with Fermat's notes. It was a tantalizing problem because of its comprehensibility. Anyone with a high-school education in mathematics can understand the statement of the problem, and many mathematicians dreamed of solving it when they were young. Despite the efforts of hundreds of amateurs and prizes offered for the solution, no correct proof was found for more than 350 years. On June 23, 1993, the British mathematician Andrew Wiles (b. 1953) announced at a conference held at Cambridge University that he had succeeded in proving a certain conjecture in algebraic geometry known as the Shimura–Taniyama conjecture, from which Fermat's conjecture is known to follow. This was the first claim of a proof by a reputable mathematician using a technique that is known to be feasible, and the result was tentatively endorsed by other mathematicians of high reputation. After several months of checking, some doubts arose. Wiles had claimed in his announcement that certain techniques involving what are called Euler systems could be extended in a particular way, and this extension proved to be doubtful. In collaboration with another British mathematician, Richard Taylor, Wiles eventually found an alternative approach that simplified the proof considerably, and there is now no doubt among the experts in number theory that the problem has been solved.

To give another illustration of the same method, we consider the problem following the one just discussed, that is, Problem 9 of Book II: *Separate a given number that is the sum of two squares into two other squares*. (That is, given one representation of a number as a sum of two squares, find a new representation of the same type.) Diophantus shows how to do this using the example $13 = 2^2 + 3^2$. He lets one of the two squares be $(\zeta + 2)^2$ and the other $(2\zeta - 3)^2$, resulting in the equation $5\zeta^2 - 8\zeta = 0$. Thus, $\zeta = \frac{8}{5}$, and indeed $(\frac{18}{5})^2 + (\frac{1}{5})^2 = 13$. It is easy to see here that Diophantus is deliberately choosing a form for the solution that will cause the constant term to drop out. This amounts to a general method, used throughout the first two books, and based on the proportion

$$(a + Y) : X = X : (a - Y)$$

for solving the equation $X^2 + Y^2 = a^2$.

The method Diophantus used to solve such problems in his first two books was conjectured by Maximus Planudes (1255–1305) and has recently been explained in simple language by Christianidis (1998).

Some of Diophantus' indeterminate problems reach a high degree of complexity. For example, Problem 19 of Book 3 asks for four numbers such that if any of the numbers is added to or subtracted from the square of the sum of the numbers, the result is a square number. Diophantus gives the solutions as

$$\frac{17, 136, 600}{163, 021, 824}, \frac{12, 675, 000}{163, 021, 824}, \frac{15, 615, 600}{163, 021, 824}, \frac{8, 517, 600}{163, 021, 824}.$$

PROBLEMS AND QUESTIONS

Mathematical Problems

- 9.1.** Use the fact that the greatest common divisor of 26173996849 and 180569389 is 713713 to reduce the fraction $\frac{180569389}{26173996849}$ to lowest terms.
- 9.2.** The Euclidean algorithm focuses on the remainders in the division process and has nothing to say about the quotients. Note that the quotients in the example given above were (in order of division) 144, 1, 20, and 12. Consider the continued fraction

$$144 + \frac{1}{1 + \frac{1}{20 + \frac{1}{12}}}.$$

Evaluate this fraction, and compare it with the result of the preceding exercise. We shall have further use for the quotients in the Euclidean algorithm when we study the mathematics of the Hindus.

- 9.3.** Verify that

$$27^5 + 84^5 + 110^5 + 133^5 = 144^5.$$

See L. J. Lander and T. R. Parkin, “Counterexample to Euler’s conjecture on sums of like powers,” *Bulletin of the American Mathematical Society*, **72** (1966), p. 1079. Smaller counterexamples to this conjecture have been discovered more recently.

Historical Questions

- 9.4. What are the main topics investigated in ancient Greek number theory?
- 9.5. What mathematical ideas were ascribed to the Pythagoreans by ancient commentators?
- 9.6. What innovation in mathematics arises in the number-theoretic investigations of Diophantus?

Questions for Reflection

- 9.7. For what reason would the ancient Greeks have been investigating figurate numbers, perfect numbers, and the like? Did they have a practical application for these ideas?
- 9.8. How much of number theory has a practical application nowadays? (If you don’t know about RSA codes, for example, look them up on-line.)
- 9.9. Should the work of Diophantus be classified as primarily number theory or primarily algebra?

Fifth-Century Greek Geometry

It is easy to surmise what problems the pre-Euclidean geometers must have worked on. One has only to look at the propositions in Euclid's *Elements*, which was, as its name implies, an *elementary* textbook of geometry and number theory, summarizing in systematic fashion what had gone before. It was certainly not the most advanced mathematics of its day, since its basic geometric tools are lines (what we now call line segments), circles, planes, and spheres. The conic sections, which were known before Euclid, and on which Euclid himself wrote a treatise, are not mentioned in it. As we have said, later commentators ascribed this geometry to the Pythagoreans. The historical problem is to trace a line of development from the basic facts Thales is said to have known to the elaborate systematic treatise of Euclid three centuries later. For guidance, we have the statements of the commentators, but they provide only a few points of light. To get a more comprehensive picture, we need to use our imaginations and conjecture one. It may not be correct, but at least it provides some coherence to the narrative and can be modified or rejected if it is incompatible with hard historical facts. The reader is hereby warned that we are about to write such a scenario and will therefore adopt a skeptical attitude toward it.

10.1. "PYTHAGOREAN" GEOMETRY

Proclus mentions two topics of geometry as being Pythagorean in origin. One is the theorem that the sum of the angles of a triangle is two right angles (Book 1, Proposition 32). Since this statement is equivalent to Euclid's parallel postulate, it is not clear what the discovery amounted to or how it was made.

10.1.1. Transformation and Application of Areas

The other topic mentioned by Proclus is a portion of Euclid's Book 6 that is not generally taught any more, called application of areas.

That topic had to be preceded by the simpler topic of transformation of areas. In his *Nine Symposium Books*,¹ Plutarch called the transformation of areas "one of the most geometrical" problems. He thought solving it was a greater achievement than discovering

¹The book is commonly known as *Convivial Questions*. The Greek word *sympósiōn* means literally *drinking together*.

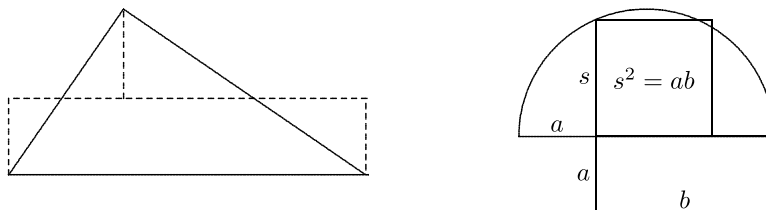


Figure 10.1. Left: turning a triangle into a rectangle. Right: turning a rectangle into a square ($s^2 = ab$).

the Pythagorean theorem and said that Pythagoras was led to make a sacrifice when he solved the problem. The basic idea is to convert a figure having one shape to another shape while preserving its area, as in Fig. 10.1. To describe the problem in a different way: Given two geometric figures A and B , construct a third figure C the same size as A and the same shape as B . One can imagine many reasons why this problem would be attractive. If one could find, for example, a square equal to any given figure, then comparing sizes would be simple, merely a matter of converting all areas into squares and comparing the lengths of their sides. But why stop at that point? Why not consider the general problem of converting any shape into any other? For polygons this problem was solved very early, and the solution appears in Proposition 25 of Euclid’s Book 6, which shows how to construct a polygon of prescribed shape equal in area to another polygon of possibly different shape.

The problem of application of areas is one degree more complicated than simply transforming an area from one shape to another. There are two such problems, both involving a given straight line segment AB and a planar polygon Γ . The first problem is to construct a parallelogram equal to Γ on part of the line segment AB in such a way that the parallelogram needed to fill up a parallelogram on the entire base, called the *defect*, will have a prescribed shape. This is the problem of *application with defect*, and the solution is given in Proposition 28 of Book 6. The second application problem is to construct a parallelogram equal to Γ on a base that is an extension of the line AB in such a way that the portion of the parallelogram extending beyond AB (the *excess*) will have a prescribed shape. This is the problem of *application with excess*, and the solution is Proposition 29 of Book 6.

The construction for application with defect is shown in Fig. 10.2. This problem does not have a solution for all given lines and areas, since the largest parallelogram that can

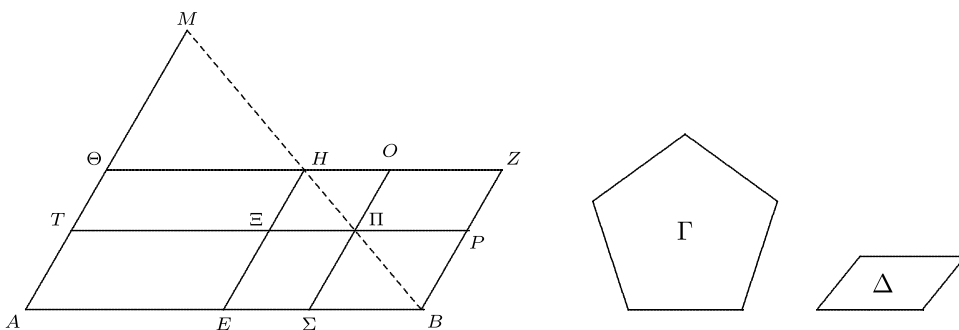


Figure 10.2. Application with defect. Euclid, Book 6, Proposition 28. Line AB , plane region Γ , and parallelogram Δ are given. Then parallelogram $A\Sigma\Pi T$ is constructed on part of line AB so as to be equal to Γ , while the *defect* $\Sigma B P \Pi$ is similar to Δ .

be formed under these conditions is the one whose base is half of the given line (Book 6, Proposition 26). Assuming that the given polygon Γ is smaller than this parallelogram, let AB be the given line, Γ the given polygonal region, and Δ the given parallelogram shape. The dashed line from B makes the same angle with AB that the diagonal of the parallelogram Δ makes with its base. The line AM is drawn to make the same angle as the corresponding sides of Δ . Then any parallelogram having its sides along AB and AM and opposite corner Π from A on the dashed line will automatically generate a "defect" that is similar to Δ . The remaining problem is to choose Π so that $A\Sigma\Pi T$ has the same area as Γ . That is achieved by constructing the parallelogram $H\Xi\Pi O$ similar to Δ and equal to the difference between $AEH\Theta$, where H is the midpoint of MB and Γ . Constructing $H\Xi\Pi O$ is the simpler transformation-of-area problem.

Besides these two problems, there is a much simpler problem of pure application, that is, finding the proper altitude for a parallelogram on the base AB so that the area is Γ . The Greek word for application is *parabolē*. Proclus cites Eudemus in asserting that the solution of the application problems was an ancient discovery of the Pythagoreans and that they gave these problems the names *ellipse* (application with defect) and *hyperbola* (application with excess), names that were later transferred to the conic curves by Apollonius. This version of events was also reported by Pappus. We shall see the reason for the transfer below.

Although most of Euclid's theorems have obvious interest from the point of view of anyone curious about the world, the application problems raise a small mystery. Why were the Pythagoreans interested in them? Were they merely a refinement of the transformation problems? Why would anyone be interested in applying an area so as to have a defect or excess of a certain shape? Without restriction on the shape of the defect or excess, the application problem does not have a unique solution. Were the additional conditions imposed simply to make the problem determinate? Some historians have speculated that there was a further motive.

In the particular case when the excess or defect is a square, these problems amount to finding two unknown lengths given their sum and product (application with defect) or given their difference and product (application with excess). In modern terms, these two problems amount to quadratic equations. (Pure application amounts to a linear equation.) Several prominent historians in the mid-twentieth century endorsed the view that Euclid's Book 2 was merely a translation into geometric language of the computational techniques found on the cuneiform tablets. And indeed, both do correspond mathematically to what we *nowadays* write as linear and quadratic equations. But neither the cuneiform writers nor Euclid had any concept corresponding to our word *equation*. Therefore neither of them was doing algebra as we understand it, and there is no reason to think that the Greek geometers were translating Mesopotamian techniques into geometric language. This hypothesis of "geometric algebra" was severely attacked by Unguru (1975/76), and no longer has many defenders.

Some historians have argued that this "geometric algebra" was a natural response to the discovery of incommensurable magnitudes, which will be discussed below, indeed a logically necessary response. On this point, however, many others disagree. Gray, for example, says that, while the discovery of incommensurables did point out a contradiction in a naive approach to ratios, "it did not provoke a foundational crisis." Nor did it force the Pythagoreans to recast algebra as geometry. In fact, it is premature to speak of equations or algebra in connection with the Greeks at this point. They had figurate numbers, among which were square numbers. At the most, we can admit that they may have looked for the side of a square equal to a certain multiple of another square. Such a problem can be considered without thinking about equations at all. Gray (1989, p. 16) concludes that

“[r]ather than turning from algebra to geometry, . . . the Greeks were already committed to geometry.”

The problem of incommensurables just mentioned was one of three challenges that one can easily imagine the early geometers having to face, once they set off down the road of a systematic, logical development of the subject to replace the isolated results achieved during the earlier period in which the main problem was to get a numerical value for an area or volume. We shall see that one of these three challenges could be ignored as far as mathematics itself was concerned, but the other two were genuine stimuli to further work and proved very fruitful, extending the Euclidean approach, which was based on two- and three-dimensional figures generated by straight lines and circles, to the limits of its potential. (After that, except for the revival of some metrical methods that had not formed part of the Euclidean canon, Greek geometry declined for lack of new material.) Let us now look at these three problems as they may have arisen. In this chapter, we shall merely state the problems. The partial solutions found to them will form most of the subject matter of the next chapter.

10.2. CHALLENGE NO. 1: UNSOLVED PROBLEMS

Supposing that the techniques of transformation and application of areas were known to the fifth-century geometers, we can easily guess what problems they would have been trying to solve. There are three natural directions in which the plane geometry of lines and circles could be extended.

1. Having learned how to convert any polygon to a square of equal area, any geometer would naturally want to do the same with circles and sectors and segments of circles. This problem was known as *quadrature (squaring) of the circle*.
2. Having solved the transformation problems for a plane, one would want to solve the analogous problems for solid figures—in other words, convert a polyhedron to a cube of equal volume. Finding the cube would be interpreted as finding the length of its side. Now, the secret of solving the planar problem was to triangulate a polygon, construct a square equal to each triangle, then add the squares to get bigger squares using the Pythagorean theorem. By analogy, the three-dimensional program would be to cut a polyhedron into tetrahedra, convert any tetrahedron into a cube equal to it, and then find a way of adding cubes analogous to the Pythagorean theorem for adding squares. The natural first step of this program (as we imagine it to have been) was to construct a cube equal to the double of a given cube, the problem of *doubling the cube*, just as we conjectured in Chapter 5 that doubling a square may have led to the Pythagorean theorem.
3. The final extension of plane geometry is the problem of dividing an arc (or angle) into equal parts. If we suppose that the fifth-century geometers knew how to bisect arcs (Proposition 9 of Book 1 of the *Elements*) and how to divide a line into any number of equal parts (Proposition 9 of Book 6), this asymmetry between their two basic figures—lines and circles—would very likely have been regarded as a challenge. The first step in this problem would have been to divide any circular arc into three equal parts, the problem of *trisection of the angle*.

The three problems just listed were mentioned by later commentators as an important challenge to all geometers. To solve them, geometers had to enlarge their set of basic objects beyond lines and planes. They were rather conservative in doing so, first invoking familiar surfaces such as cones and cylinders, which could be generated by moving lines on circles, and intersecting them with planes so as to get the conic sections that we know as the ellipse, parabola, and hyperbola. These curves made it possible to solve two of the three problems (trisecting the angle and doubling the cube). Later, a number of more sophisticated curves were invented, among them spirals, and the quadratrix. This last curve got its name from its use in squaring the circle. Although it is not certain that the fifth-century geometers had a program like the one described above, it is known that all three of these problems were worked on in antiquity. Solving these problems was certainly a desirable goal, but that solution could take its time. Mathematical problems only become more interesting when they remain unsolved for an extended period. Not solving them in no way threatened the achievements already gained.

10.3. CHALLENGE NO. 2: THE PARADOXES OF ZENO OF ELEA

Although we have some idea of the geometric results proved by the early Greek geometers, our knowledge of their interpretation of these results is murkier. How did they conceive of geometric entities such as points, lines, planes, and solids? Were these objects physically real or merely ideas? What properties did they have? Some light is shed on this question by the philosophical critics, one of whom has become famous for the paradoxes he invented.

As mentioned, in the Pythagorean philosophy, the universe was said to have been generated by numbers and motion. That these concepts needed to be sharpened up became clear from critics of a naive view of geometry. We now know that the basic problem is the incompatibility between discrete modes of thought and continua. (As we shall see below, the third challenge—that of incommensurable pairs of lines—arises precisely because lines are continuous.) It turns out to be more difficult to think about continuous media than one might imagine.

These paradoxes are ascribed to the philosopher Zeno of Elea. Zeno died around 430 BCE, and we do not have any of his works to rely on, only expositions of them by other writers. Aristotle, in particular, says that Zeno gave four puzzles about motion, which he called the Dichotomy (division), the Achilles, the Arrow, and the Stadium. Here is a summary of these arguments in modern language, based on Book 6 of Aristotle's *Physics*.

1. *The Dichotomy*. Motion is impossible because before an object can arrive at its destination it must first arrive at the middle of its route. Then before it can arrive at the end, it must reach the midpoint of the second half of the route, and so forth. Thus we see that the object must do infinitely many things in a finite time in order to move.
2. *The Achilles*. (This paradox is apparently so named because in Homer's *Iliad* the legendary warrior Achilles chased the Trojan hero Hector around the walls of Troy, overtook him, and killed him.) If given a head start, the slower runner will never be overtaken by the faster runner. Before the two runners can be at the same point at the same instant, the faster runner must first reach the point from which the slower runner started. But at that instant the slower runner will have reached another point ahead of the faster. Hence the race can be thought of as beginning again at that instant, with the slower runner still having a head start. The race will "begin again" in this sense

infinitely many times, with the slower runner always having a head start. Thus, as in the dichotomy, infinitely many things must be accomplished in a finite time in order for the faster runner to overtake the slower.

3. *The Arrow*. An arrow in flight is at rest at each instant of time. That is, it does not move from one place to another during that instant. But then it follows that it cannot traverse any positive distance because successive additions of zero will never result in anything but zero.
4. *The Stadium*. (In athletic stadiums in Greece the athletes ran from the goal, around a halfway post and then back. This paradox seems to have been inspired by imagining two lines of athletes running in opposite directions and meeting each other.) Consider two parallel line segments of equal length moving in opposite directions with equal speeds and a third line that is stationary and located between the two of them. The speed of each line is measured by the number of points of space it passes over in a given time. In the time required for a point of each line to pass a point of the other, these two points apparently pass only half of a point on the stationary line. Since there is no such thing as half a point, it appears that the speed of each line relative to the stationary line must be twice what it appears to be.

Even today, we think of a line as “made of” points, but Zeno’s paradoxes seem to show that space cannot be “made of” points in the same way that a building can be made of bricks. For assuredly the number of points in a line segment cannot be finite. If it were, since points are indivisible (*atoms* in the original Greek sense of the word), the line would not be infinitely divisible as the dichotomy and Achilles paradoxes showed that it must be; moreover, the stadium paradox would show that the number of points in a line equals its double. There must therefore be an infinity of points in a line. But then each of these points must take up no space; for if each point occupied some space, an infinite number of them would occupy an infinite length.² But if points occupy no space, how can the arrow, whose tip is at a single point at each instant of time, move through a *positive* quantity of space? A continuum whose elements are points was needed for geometry, yet it could not be thought of as being made up of points in the way that discrete collections are made up of individuals.

This challenge, while it no doubt provided brain-breaking puzzles for mathematicians for a long time, can nevertheless be ignored by those who have no metaphysical bent and are concerned only with deriving one statement from another by logical deduction. Geometers had no acute need to solve the problems posed by Zeno, even though they pointed up difficulties with the *interpretation* of mathematical concepts. Like the unsolved construction problems listed above, leaving them unanswered posed no threat to the formal creation of geometry.

10.4. CHALLENGE NO. 3: IRRATIONAL NUMBERS AND INCOMMENSURABLE LINES

The difficulties pointed out by Zeno affected the intuitive side of geometry and its interpretation. We would call them metaphysical puzzles rather than mathematical puzzles nowadays.

²Keep in mind that a line, to the Greeks, was what we now call a line segment. It was not infinitely long.

The challenge they posed, which involved elucidating the nature of a continuum, was not satisfactorily met until the late nineteenth and early twentieth century. (Some say not even then!) There was, however, a challenge that came from within the formal system of geometry. To the modern mathematician, this second challenge in dealing with the concept of a continuum is much more pertinent and interesting than the paradoxes of Zeno. That challenge is the problem of incommensurables, which led ultimately to the concept of a real number.

The existence of incommensurables throws doubt on certain oversimplified proofs of proportion. When two lines or areas are commensurable, one can describe their ratio as, say, $5 : 7$, meaning that there is a common measure such that the first object is five times this measure and the second is seven times it. A proportion such as $a : b :: c : d$, then, is the statement that ratios $a : b$ and $c : d$ are both represented by the same pair of numbers. Almost certainly, the legendary aphorism of Pythagoras, that “all is number,” refers to this use of integers to define the ratio of two objects.³ It was therefore problematic when pairs of lines were discovered that had no common measure, and whose ratio could therefore not be expressed in this way. As the concept of incommensurable pairs of lines is intimately bound up with what we now call *irrational numbers* (and were not considered numbers at all by the Greeks working in the Euclidean tradition), we shall look at these two phenomena together and compare them.

The absence of a place-value system of writing numbers forced the Greek mathematicians to create a way around the problem that other societies have dealt with through rational approximations. Place-value notation provides approximate square roots in practical form, even when the expansion does not terminate. We already mentioned, in Chapter 5, a cuneiform tablet from Iraq (YBC 7289 from the Yale Babylonian Collection) showing a square with its diagonals drawn and the sexagesimal number $1;24,51,10$, which gives the length of the diagonal of a square of side 1 to great precision. This rational sexagesimal number surely represents the irrational “number” $\sqrt{2}$.

The word *number* is placed in inverted commas here because the meaning of the square root of 2 is not easy to define. One quickly gets into a vicious circle when trying to formulate its definition. The difficulty came in a clash of geometry and arithmetic, the two fundamental modes of mathematical thinking. From the arithmetical point of view the problem is minimal. If numbers must be what we now call positive rational numbers, then some of them are squares and some are not, just as some integers are triangular, square, pentagonal, and so forth, while others are not. No one would be disturbed by this fact. Since the Greeks had no place-value system to suggest an infinite process leading to an exact square root, they might not have speculated deeply on the implications of this arithmetical distinction in geometry. In other words, they, like their predecessors, had no reason to think about what we call infinitely precise real numbers. They did, however, speculate on both the numerical and geometric aspects of the problem, as we shall now see.

Just when the problem of irrationals and incommensurables arose cannot be specified very exactly. Probably it was in the early fourth century BCE, and certainly before 350 BCE. Since the problem has both numerical and geometric aspects, we begin with the numerical problem.

³Since no writings of Pythagoras or his immediate followers survive, it is not possible to find this aphorism stated so concisely anywhere. In his *Metaphysics*, Bekker 985b, Aristotle says that the Pythagoreans “supposed the elements of numbers to be the elements of all things.”

10.4.1. The Arithmetical Origin of Irrationals

In Plato's dialogue *Theatetus*, the title character reports that a certain Theodorus proved that the integers 2, 3, 5, and so on, up to 17 have no (rational) square roots, except of course the obvious integers 1, 4, and 9; and he says that for some reason, Theodorus got stuck at that point. On that basis the students decided to classify numbers as *equilateral* and *oblong*. The former class consists of the squares of rational numbers, for example $\frac{25}{9}$, and the latter are all other positive rational numbers, such as $\frac{3}{2}$, which cannot be written as a product of two equal factors.

One cannot help wondering why Theodorus got stuck at 17 after proving that the numbers 3, 5, 6, 7, 8, 10, 11, 12, 13, 14, and 15 have no square roots. What might the difficulty have been? The square root of 17 is irrational, and the proof commonly given nowadays to show the irrationality of $\sqrt{3}$, for example, based on the unique prime factorization of integers, works just as well for 17 as for any other nonsquare integer. If Theodorus had our proof, he wouldn't have gotten stuck doing 17, and he wouldn't have bothered to do so many special cases, since the proofs are all the same. Therefore, we must assume that he was using some other method.

An ingenious conjecture as to Theodorus' method was provided by the late Wilbur Knorr (1945–1997) in his book (1975). Knorr suggested that the proof was based on the elementary distinction between even and odd. To see how such a proof works, suppose that 7 is an equilateral number in the sense mentioned by *Theatetus*. Then there must exist two integers m and n such that $m^2 = 7n^2$. We can assume that both integers are odd, since if both are even, we can divide them both by 2, and it is impossible for one of them to be odd and the other even. (The fact that the square of one of them equals seven times the square of the other would imply that an odd integer equals an even integer if this were the case.) Now it is well known that the square of an odd integer is always 1 larger than a multiple of 8. The supposition that the one square is seven times the other then implies that an integer 1 larger than a multiple of 8 equals an integer 7 larger than a multiple of 8, which is clearly impossible.

This same argument shows that none of the odd numbers 3, 5, 7, 11, 13, and 15 can be the square of a rational number. With a slight modification, it can also be made to show that none of the numbers 2, 6, 8, 10, 12, and 14 is the square of a rational number, although no argument is needed in the case of 8 and 12, since it is already known that $\sqrt{2}$ and $\sqrt{3}$ are irrational. Notice that the argument fails, as it must, for 9: A number 9 larger than a multiple of 8 is also 1 larger than a multiple of 8. However, it also breaks down for 17 and for the same reason: A number 17 larger than a multiple of 8 is also 1 larger than a multiple of 8. Thus, even though it is *true* that 17 is not the square of a rational number, the argument just given, based on what we would call arithmetic modulo 8, cannot be used to *prove* this fact. In this way the conjectured method of proof would explain why Theodorus got stuck at 17.

Theodorus thus proved not only that there was no *integer* whose square is, say, 11 (which is a simple matter of ruling out the few possible candidates), but also that there was not even any *rational number* having this property; that is, 11 is not the square of anything the Greeks recognized as a number.

10.4.2. The Geometric Origin of Irrationals

A second, "geometric" theory of the origin of irrational numbers comes from geometry and seems less plausible. If we apply the Euclidean algorithm to the side and diagonal of

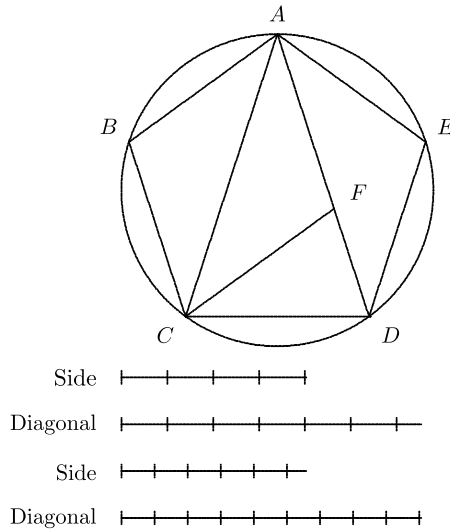


Figure 10.3. Diagonal and side of a regular pentagon. If a unit is chosen that divides the side into equal parts, it cannot divide the diagonal into equal parts, and vice versa.

the regular pentagon in Fig. 10.3, we find that the diagonal AD and the CD get replaced by lines equal, respectively, to CD (which equals CF , the bisector of angle ACD , which in turn equals AF) and DF , and these are the diagonal and side of a smaller pentagon since $\angle ACD = 2\angle FCD$. Thus, no matter how many times we apply the procedure of the Euclidean algorithm, the result will always be a pair consisting of the side and diagonal of a pentagon. Therefore, in this case the Euclidean algorithm will *never* produce an equal pair of lines. We know, however, that it *must* produce an equal pair if a common measure exists. We conclude that *no common measure can exist for the side and diagonal of a pentagon*. The same is true for the side and diagonal of a square, although the algorithm requires two applications in order to cycle. The absence of a common measure for the side and diagonal of a square is the exact geometric equivalent of the arithmetic fact that there is no rational number whose square is 2. In other words, incommensurable magnitudes and irrational numbers (as we think of them—again, they were not numbers to the Greeks) are two different ways of looking at the same phenomenon.

The argument just presented was originally given by von Fritz (1945). Knorr (1975, pp. 22–36) argued against this approach, however, pointing out that the simple arithmetic relation $d^2 = 2s^2$ satisfied by the diagonal and side of a square can be used in several ways to show that d and s could not both be integers, no matter what length is chosen as unit. Knorr preferred a reconstruction closer to the argument given in Plato's *Meno*, in which the problem of doubling a square is discussed. Knorr pointed out that when discussing irrationals, Plato and Aristotle always invoke the side and diagonal of a square, never the pentagon or the related problem of dividing a line in mean and extreme ratio, which they certainly knew about.

10.4.3. Consequences of the Discovery

Whatever the argument used may have been, the Greeks somehow discovered the existence of incommensurable pairs of line segments before the time of Plato. If indeed Pythagorean

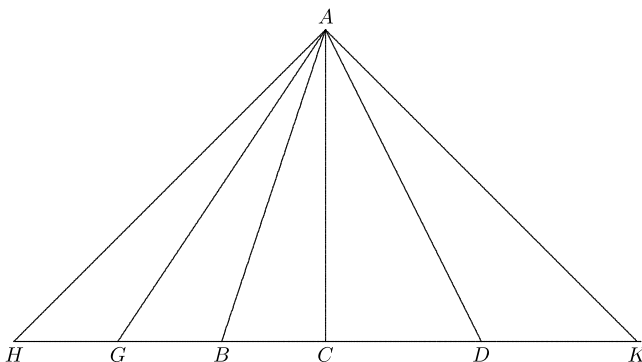


Figure 10.4. A fundamental theorem in the theory of proportion. Proposition 1 of Book 6 of the *Elements*.

metaphysics was what it appears to be, this discovery must have been disturbing: Number, it seems, is *not* adequate to explain all of nature. As mentioned in Chapter 8, a legend arose that the Pythagoreans attempted to keep secret the discovery of this paradox. However, scholars believe that the discovery of incommensurables came near the end of the fifth century BCE, when the original Pythagorean group was already defunct.

The existence of incommensurables throws doubt on certain oversimplified proofs of geometric proportion, as we shall now show. This theory of proportion is extremely important in geometry if we are to have such theorems as Proposition 2 of Book 12 of Euclid's *Elements*, which says that circles are proportional (in area) to the squares on their diameters. Even the simplest constructions, such as the construction of a square equal in area to a given rectangle or the application problems mentioned above, may require the concept of proportionality of lines. Because of the importance of the theory of proportion for geometry, the discovery of incommensurables made it imperative to give a definition of proportion without relying on a common measure to define a ratio.

To see why the discovery of incommensurables created a problem, although perhaps not a scandal, consider the following conjectured early proof of a fundamental result in the theory of proportion: the proposition that two triangles having equal altitudes have areas proportional to their bases. This assertion is half of Proposition 1 of Book 6 of Euclid's *Elements*. Let ABC and ACD in Fig. 10.4 be two triangles having the same altitude. Euclid draws them as having a common side, but that is only for convenience. This positioning causes no loss in generality because of the proposition that any two triangles of equal altitude and equal base are equal, proved as Proposition 38 of Book 1.

Suppose that the ratio of the bases $BC : CD$ is $2 : 3$, that is, $3BC = 2CD$. Extend BD leftward to H so that $BC = BG = GH$, producing triangle AHC , which is three times triangle ABC . Then extend CD rightward to K so that $CD = DK$, yielding triangle ACK equal to two times triangle ACD . But then, since $GC = 3BC = 2CD = CK$, triangles AGC and ACK are equal. Since $AGC = 3ABC$ and $ACK = 2ACD$, it follows that $ABC : ACD = 2 : 3$. We, like Euclid, have no way of actually *drawing* an unspecified number of copies of a line, and so we are forced to *illustrate* the argument using specific numbers (2 and 3 in the present case) while expecting the reader to understand that the argument is completely general.

An alternative proof could be achieved by finding a common measure of BC and CD , namely $\frac{1}{2}BC = \frac{1}{3}CD$. Then, dividing the two bases into parts of this length, one would have divided ABC into two triangles and divided ACD into three triangles, and all five of the smaller triangles would be equal. But both of these arguments fail if no integers m and n can be found such that $mBC = nCD$, or (equivalently) no common measure of BC and CD exists. This proof needs to be shored up, but how is that to be done? We shall see in the next chapter.

Like Gray (quoted above), Knorr (1975) argued that the discovery of irrationals was not a major “scandal,” and that it was not responsible for the “geometric algebra” in Book 2 of Euclid. While arguing that incommensurability forced some modifications in the way of thinking about physical magnitudes, he said (p. 41):

It is thus thoroughly obvious that, far from being in a state of paralysis, fifth- and fourth-century geometers proceeded with their studies of similar figures as if they were still unaware of the foundational consequences of the existence of incommensurable lines.

PROBLEMS AND QUESTIONS

Mathematical Problems

- 10.1.** The problem of application with defect discussed above requires constructing a parallelogram of a given shape that is equal to the difference of two other parallelograms of the same shape. Show how to do this using the Pythagorean theorem. (Assume the given shape is square, if it makes the problem easier. In fact, as we shall see in Chapter 13, Euclid shows in Book VI of the *Elements* that the Pythagorean theorem works just as well for similar parallelograms as it does for squares. It should be noted that Euclid elegantly shortens this construction, as he so often does.)
- 10.2.** Assuming that there are two square integers whose ratio is 5, derive a contradiction using the principle that underlies Knorr’s conjecture. (If the integers are relatively prime, then both must be odd. Use that fact and the fact that the square of any odd number is one unit larger than a multiple of 8 to derive a contradiction.)
- 10.3.** The ratio of the diagonal of a pentagon to its side has been called the Golden Ratio for many centuries. It is usually denoted Φ , and its exact value is $\frac{1+\sqrt{5}}{2}$. Use the fact that $\Phi = 1 + \frac{1}{\Phi}$ to show that the Euclidean algorithm applied to find a common measure of Φ and 1 will go on forever producing a quotient of 1, but never terminate. (Keep substituting the entire right-hand side of this equation for the Φ that occurs there in the denominator.)

Historical Questions

- 10.4.** What achievements do Proclus and Plutarch ascribe to the Pythagoreans?
- 10.5.** Summarize the four Zeno paradoxes, as reported by Aristotle.
- 10.6.** What were the three classical problems of geometry worked on during the fifth and fourth centuries BCE?

Questions for Reflection

- 10.7. What motive could the early Greek geometers have had for studying the problems of transformation and application of areas?
- 10.8. How do you know that there “exists” a number whose square is 2? In what sense do we know what this number is?
- 10.9. How do you resolve the paradoxes of Zeno?

Athenian Mathematics I: The Classical Problems

The fifth century BCE was the high-water mark of Athenian power. The Ionian islands were constantly menaced and often subjugated by the Persian Empire, which also occasionally threatened the mainland of Greece. In 490 BCE, the Athenians stood alone and fought off a Persian invasion at the Battle of Marathon, thereby increasing their prestige among the Greek city-states. Ten years later, when another invasion was imminent, the famous 300 Spartans commanded by Leonidas, along with about a thousand others, held off the Persians for several days before being overwhelmed by the superior numbers of the Persian army at Thermopylae. This victory allowed the Persians to invade Greece and sack Athens, but the Greek naval forces led by the Athenians defeated the Persian navy at Salamis, forcing the Persians to delay the conquest of the rest of Greece. The following year, they were defeated by a combined Greek force at Plataea. Once the Persian threat was beaten back, the Spartans withdrew into isolationism, while the Athenians vigorously promoted a Greek defense league, with themselves at the head of it. Athens became quite prosperous during the period of peace. Some of the magnificent buildings whose ruins still inspire the visitor were built during the time of Pericles' leadership of the city (461–429). It was during this time that the philosopher Anaxagoras (ca. 500–428) came to Athens and eventually was arrested on the charge of denying that the sun was the god Helios. (He taught that it was a hot stone the size of the Peloponnesus.) While in prison, he allegedly worked on the problem of squaring the circle.

In 431 BCE, war broke out between Sparta and Athens and raged intermittently for the next quarter-century. Fortune seemed to turn against the Athenians on nearly every occasion, and finally, in 404 BCE, they capitulated. The Spartans had no desire to colonize or rule Athens; and they quickly restored the Athenian government, which proceeded to take revenge on the aristocrats who had sided or appeared to side with the Spartans. Among the victims was the philosopher Socrates (ca. 470–399), one of whose followers was Plato, also a member of an aristocratic family. After Socrates' death, Plato journeyed to Sicily, where he is said to have met the Pythagorean philosopher Philolaus. Returning to Athens, he founded his famous Academy in 387 BCE.

Plato's main interest, conditioned no doubt by his experience of war and defeat, was in political questions. He wished to get the very best people to rule. Among the virtues, he placed a high value on knowledge and wisdom, and through that route he became interested

in mathematical questions. Some of his students worked on mathematical problems. Because of this shift in the intellectual center of gravity from the commercial Greek colonies to Athens, we are going to call the geometry developed during the late fourth century BCE and throughout the third century *Athenian mathematics*, even though not all of it was done in Athens. One of its highest achievements, the solution of the problem of incommensurables by Eudoxus, was the work of a former disciple of Plato who had moved on and established himself elsewhere as a prominent geometer and astronomer.

Plato's most famous student, Aristotle (384–322), left the Academy just before the death of Plato and set up his own school, the Lyceum, over the hill in Athens from the Academy. His most famous pupil was Alexander (son of Philip of Macedon), later to be known as Alexander the Great.¹ The Macedonians conquered the Greek mainland and expanded their control over the entire Middle East, crushing the Persian Empire at the Battle of Arbela in 331 BCE. After conquering Egypt, Alexander founded a new city in the Nile Delta, naming it after himself. In that city, his general Ptolemy Soter, who succeeded him as ruler of that portion of the Macedonian Empire, founded the famous Library, at which the great mathematicians of the third century BCE all studied.

Let us now proceed to examine this work that we are calling Athenian mathematics. In the present chapter, we shall discuss only the progress made on the three unsolved classical problems mentioned in the preceding chapter. The all-important work on the theory of incommensurables, and the logical ordering of the material will be discussed in the next chapter, which is devoted to the schools of Plato and Aristotle.

The problems of doubling the cube and trisecting the angle were solved, to the extent that they can be solved, during this period. Even so, new methods of solving them continued to be sought long afterward; and the quadrature of the circle, a much more difficult problem, was never solved in a satisfactory way. In order to tell as full a story as possible, we shall extend our discussion of these three classical problems beyond the period that we have characterized as Athenian mathematics.

11.1. SQUARING THE CIRCLE

Proclus mentions Hippocrates of Chios as having discovered the quadratures of lunes. This mathematician (ca. 470–ca. 410 BCE), who lived in Athens at the time of the Peloponnesian War, is said to have worked on all three of the classical problems. A lune is a figure resembling a crescent moon: the region inside one of two intersecting circles and outside the other. In the ninth volume of his commentary on Aristotle's books on physics, the sixth-century commentator Simplicius discusses several lunes that Hippocrates squared, including the one depicted in Fig. 11.1. After detailing the criticism by Eudemus of earlier attempts by the Sophist Antiphon (480–411) to square the circle by polygonal approximation, Simplicius reports the quadrature shown in (Fig. 11.1), which uses a result that later appeared in Book 12 of Euclid's *Elements*. The result needed is that semicircles are proportional to the squares

¹Alexander was a man of action, and there is no evidence anywhere in his entire career that Aristotle had had the slightest influence on him. He was also apparently tutored by Menaechmus (ca. 380–ca. 320), another student of Plato. The fifth-century CE writer Stobaeus writes that Alexander insisted on getting an abridged course in geometry, but Menaechmus told him there were no “kingshighways” in geometry and that everyone had to follow the same road. (Stobaeus, Book 2, Chapt. 31, §115. Proclus tells the same story with Ptolemy Soter in place of Alexander and Euclid in place of Menaechmus.)

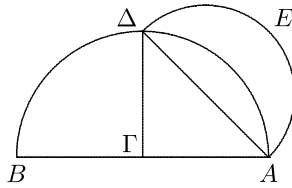


Figure 11.1. Hippocrates' quadrature of a lune, according to Simplicius.

on their diameters. How that fact was established will be taken up in the next chapter. In the meantime, we note that if $r = \overline{A\Gamma}$ is the radius of the large semicircle $A\Gamma B\Delta$, then the radius of the smaller semicircle $A\Delta E$ is $\frac{r}{\sqrt{2}}$. The segment of the larger semicircle inside the smaller one is obviously half of the larger semicircle less the triangle $A\Gamma\Delta$. In our terms, the segment has area $\frac{\pi}{4}r^2 - \frac{1}{2}r^2$, while the area of the smaller semicircle is $\frac{\pi}{4}r^2$. Therefore the lune, which is the difference of these two figures, has area equal to the triangle. Thus, quadrature of some figures bounded by circular arcs is possible, since this lune is demonstrably equal to a figure bounded by straight lines.

Simplicius' reference to Book 12 of Euclid's *Elements* is anachronistic, since Hippocrates lived before Euclid; but it was probably well known that similar circular segments are proportional to the squares on their bases. Even that theorem is not needed here, except in the case of semicircles, and that special case is easy to derive from the theorem for whole circles. The method of Hippocrates does not achieve the quadrature of a whole circle; we can see that his procedure works because the "irrationalities" of the two circles cancel each other when the segment of the larger circle is removed from the smaller semicircle.

In his essay *On Exile*, Plutarch reports that the philosopher Anaxagoras worked on the quadrature of the circle while imprisoned in Athens. Other attempts are reported, one by Dinostratus (ca. 390–ca. 320 BCE), the brother of Menaechmus. Dinostratus is said to have used the curve called (later, no doubt) the *quadratrix*, (*squarer*), said to have been invented by Hippias of Elis (ca. 460–ca. 410 BCE) for the purpose of trisecting the angle. It is discussed below in that connection.

11.2. DOUBLING THE CUBE

Although the problem of doubling the cube fits very naturally into what we have imagined as a purely geometric program—to extend the achievements in transformation of areas into similar results in the transformation of volumes—some ancient authors gave it a more exotic origin. In *The Utility of Mathematics*, Theon of Smyrna discusses a work called *Platonicus* that he ascribes to Eratosthenes. In that work, the citizens of Delos (the island that was the headquarters of the Athenian Empire) consulted an oracle in order to be relieved of a plague, and the oracle told them to double the size of an altar.² According to Theon, Eratosthenes depicted the Delians as having turned for technical advice to Plato, who told them that the altar was not the point: The gods really wanted the Delians to learn geometry better. In his

²Plagues were apparently common in ancient Greece. One is described at the outset of Homer's *Iliad* as being due to the wrath of Apollo. Another occurs in Sophocles' *Oedipus the King*, and another decimated Athens early in the Peloponnesian War, claiming Pericles as one of its victims.

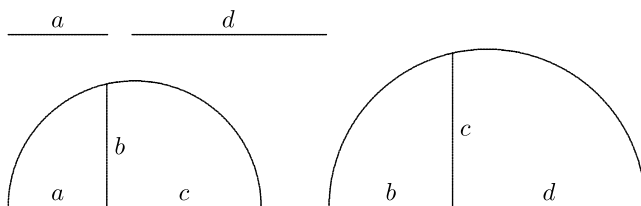


Figure 11.2. The problem of two mean proportionals: Given a and d , find b and c .

commentary on Archimedes' work on the sphere and cylinder, Eutocius gives another story, also citing Eratosthenes, but he says that Eratosthenes told King Ptolemy in a letter that the problem arose on the island of Crete when King Minos ordered that a tomb built for his son be doubled in size.

Whatever the origin of the problem, both Proclus and Eutocius agree that Hippocrates was the first to reduce it to the problem of two mean proportionals. The fifth-century geometers knew that the mean proportional between any two square integers is an integer (for example, $\sqrt{16 \cdot 49} = 28$) and that between any two cubes such as 8 and 216 there are two mean proportionals (Euclid, Book 8, Propositions 11 and 12); for example, $8 : 24 :: 24 : 72 :: 72 : 216$. If two mean proportionals could be found between the sides of two cubes—as seems possible, since every volume can be regarded as the cube on some line—the problem would be solved. It would therefore be natural for Hippocrates to think along these lines, by analogy with the result on figurate numbers, when comparing two cubes. Eutocius, however, was somewhat scornful of this reduction, saying that the new problem was just as difficult as the original one. That claim, however, is not true: One can easily draw a figure containing two lines and their mean proportional (Fig. 11.2): the two parts of the diameter on opposite sides of the endpoint of the half-chord of a circle and the half-chord itself. The only problem is to get two such figures with the half-chord and one part of the diameter reversing roles between the two figures and the other parts of the diameters equal to the two given lines, as shown in Fig. 11.2. It is natural to think of using two semicircles for this purpose and moving the chords to meet these conditions.

In his commentary on the treatise of Archimedes on the sphere and cylinder, Eutocius gives a number of solutions to this problem, ascribed to various authors, including Plato. The earliest one that he reports is due to Archytas (ca. 428–350 BCE). This solution requires intersecting a cylinder with a torus and a cone. The three surfaces intersect in a point from which the two mean proportionals can be determined. A later solution by Menaechmus may have arisen as a simplification of Archytas' rather complicated construction. It requires intersecting two cones, each having a generator parallel to a generator of the other, with a plane perpendicular to both generators. These intersections form two conic sections, a parabola and a rectangular hyperbola; where they intersect, they produce the two mean proportionals, as shown in Fig. 11.7.

If Eutocius is correct, the conic sections first appeared, but not with the names they now bear, in the fourth century BCE. Menaechmus created these sections by cutting a cone with a plane perpendicular to one of its generators. When that is done, the shape of the section depends on the apex angle of the cone. If that angle is acute, the section will be an ellipse; if it is a right angle, the section will be a parabola; if it is obtuse, the section will be a hyperbola. In the commentary on Archimedes' treatise on the sphere and cylinder mentioned above, Eutocius tells how he happened to find a work written in the Doric dialect

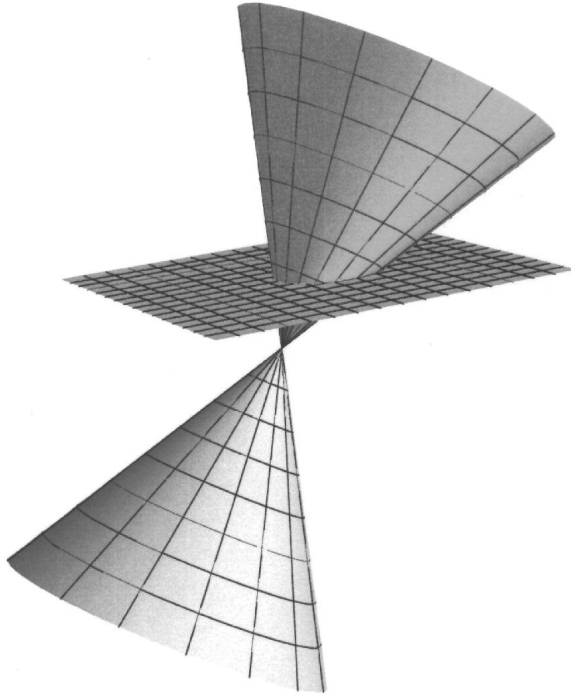


Figure 11.3. The oxytome (ellipse) of Menaechmus, obtained by cutting an acute-angled cone by a plane perpendicular to a generator.

which seemed to be a work of Archimedes. He mentions in particular that instead of the word *parabola*, used since the time of Apollonius, the author used the phrase *section of a right-angled cone*; and instead of *hyperbola*, he used the phrase *section of an obtuse-angled cone*. Since Proclus refers to “the conic section triads of Menaechmus,” it is inferred that the original names of the conic sections were *oxytomē* (sharp cut), *orthotomē* (right cut), and *amblytomē* (blunt cut), as shown in Figs. 11.3–11.5. However, Menaechmus probably thought of the cone as the portion of the figure from the vertex to some particular circular base, since the Greeks did not consider infinitely extended bodies. In particular, he wouldn’t have thought of the hyperbola as having two nappes, as we now do.

How Apollonius of Perga came to give them their modern names a century later is described below. At present, we shall look at the consequences of Menaechmus’ approach and see how it enabled him to solve the problem of two mean proportionals. It is very difficult for a modern mathematician to describe this work without breaking into modern algebraic notation, essentially using analytic geometry. It is very natural to do so, because Menaechmus, if Eutocius reports correctly, comes very close to stating his theorem in algebraic language. To describe this work in Menaechmus’ original language would require far more space than we have available and would be tedious and confusing. Thus, with apologies for the inevitable distortion, we shall abbreviate the discussion and use some algebraic symbolism.

We begin by looking at a general conic section, shown in Fig. 11.6. When a cone is cut by a plane through its axis, the resulting figure is simply a triangle, called the *axial triangle*. The end that we have left open by indicating with arrows the direction of the axis and two

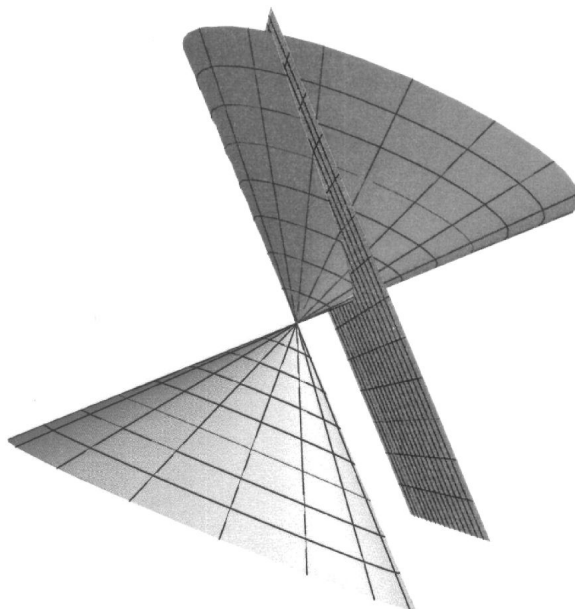


Figure 11.4. The orthotome (parabola) of Menaechmus, obtained by cutting a right-angled cone by a plane perpendicular to a generator.

generators in this plane would have been closed off by Menaechmus. If the cone is cut by a plane perpendicular to its axis, the result is a circle. The conic section is obtained as the intersection with a plane perpendicular to one of its generators at a given distance (marked u in the figure) from the apex. The important relation needed is the one between the length of a horizontal chord (double the length marked v) in the conic section and its height (marked

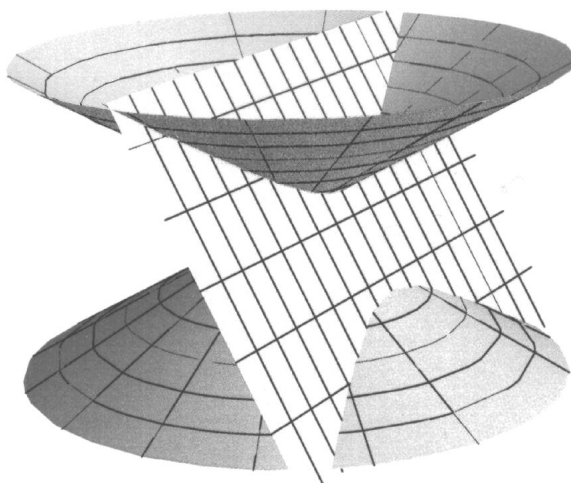


Figure 11.5. The amblytome (hyperbola) of Menaechmus, obtained by cutting an obtuse-angled cone by a plane at right angles to a generator.

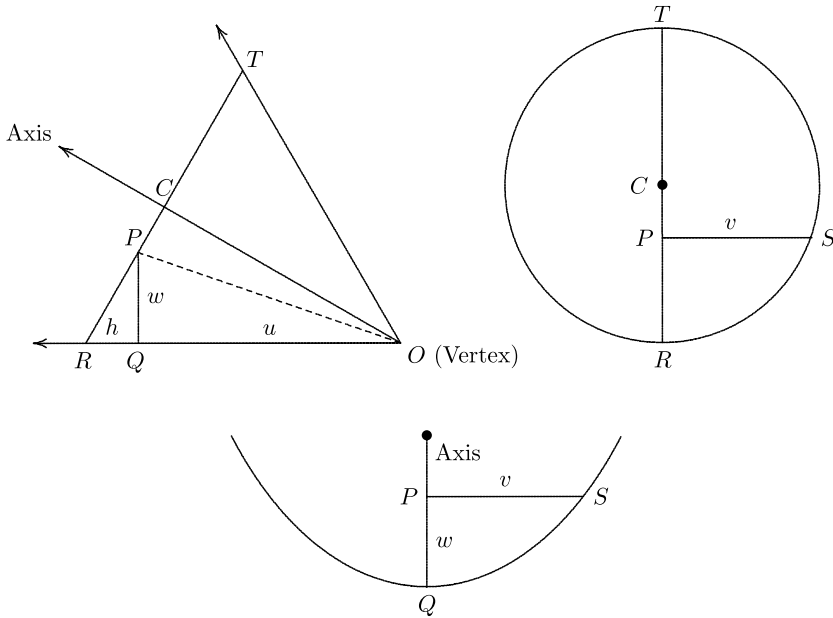


Figure 11.6. Sections of a cone. Top left: through the axis. Top right: perpendicular to the axis. Bottom: perpendicular to the generator OR at a point Q lying at distance u from the vertex O . The fundamental relation is $v^2 = h^2 + 2uh - w^2$. The length h has a fixed ratio to w , depending only on the shape of the triangle OCR .

w) above the generator that has been cut. Using only similar triangles and the fact that a half chord in a circle is the mean proportional between the segments of the diameter through its endpoint, Menaechmus derived the fundamental relation that we write as

$$v^2 = h^2 + 2uh - w^2.$$

Although we have written this relation as an equation with letters in it, Menaechmus would have been able to describe what it says in terms of the lines v , u , h , and w , and squares and rectangles on them. He would have known the value of the ratio h/w , which is determined by the shape of the triangle ROC . In our terms $h = w \tan(\varphi/2)$, where φ is the apex angle of the cone. When conic sections are to be applied, the user has free choice of the apex angle φ and the length u .

The simplest case is that of the parabola, where the apex angle is 90° and $h = w$. In that case the relation between v and w is

$$v^2 = 2uw.$$

In the problem of putting two mean proportionals B and Γ between two lines A and E , Menaechmus took the u for this parabola to be $\frac{1}{2}A$, so that $v^2 = Aw$.

The hyperbola Menaechmus needed for this problem was a rectangular hyperbola, which results when the triangle ROC is chosen so that $\overline{RC}^2 = 2\overline{OC}^2$, and therefore $\overline{OR}^2 = 3\overline{OC}^2$. Such a triangle is easily constructed by extending one side of a square to the same length as the diagonal and joining the endpoint to the opposite corner of the square. In any triangle of

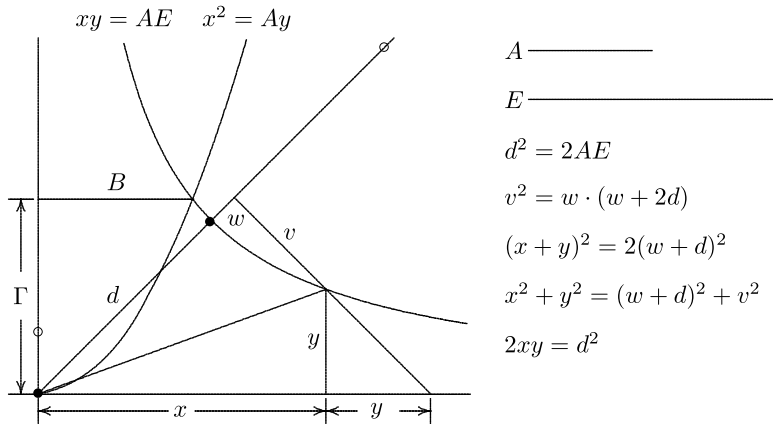


Figure 11.7. One of Menaechmus' solutions to the problem of two mean proportionals, as reported by Eutocius.

this shape the legs are the side and diagonal of a square. In this case, the apex angle of the cone is about 109.4° or 1.910634 radians, and we get $h = \sqrt{2}w$. For that case Menaechmus would have been able to show that

$$(w + \sqrt{2}u)^2 - v^2 = 2u^2,$$

that is,

$$v^2 = w(w + 2d),$$

where $d (= \sqrt{2}u)$ is the diagonal of a square whose side is u . To solve the problem of two mean proportionals, Menaechmus took $u = \sqrt{AE}$; that is, the mean proportional between A and E . Menaechmus' solution is shown in Fig. 11.7.

This solution uses only figures that arise naturally from circles and straight lines, yet people were not satisfied with it. The objection to it was that the data and the resulting figure all lie within a plane, but the construction requires the use of cones, which cannot be contained in the plane.

11.3. TRISECTING THE ANGLE

The practicality of trisecting an angle is immediately evident: It is the first step on the way to dividing a circular arc into any number of equal pieces. If a right angle can be divided into n equal pieces, a circle also can be divided into n equal pieces, and hence the regular n -gon can be constructed. Success in constructing the regular pentagon may have stimulated work on such a program. It is possible to construct the regular n -gon using only straight lines and circles for $n = 3, 4, 5, 6, 8, 10$, but not 7 or 9 . The number 7 is awkward, being the only prime between 5 and 10 , and one could expect to have difficulty constructing the regular heptagon. Surprisingly, however, the regular heptakaidecagon (17-sided polygon) can be constructed using only compass and straightedge. Since $9 = 3 \cdot 3$, it would seem natural to begin by trying to construct this figure, that is, to construct an angle of 40° . That would

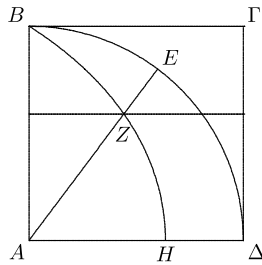


Figure 11.8. The quadratrix of Hippias.

be equivalent to constructing an angle of 20° , hence trisecting the angles of an equilateral triangle.

Despite the seeming importance of this problem, less has been written about the ancient attempts to solve it than about the other two problems. For most of the history we are indebted to two authors. In his commentary on Euclid's *Elements*, Proclus mentions the problem and says that it was solved by Nicomedes using his conchoid and by others using the quadratrices of Hippias and Nicomedes (280–210). In Book 4 of his *Synagōgē (Collection)*, Pappus says that the circle was squared using the curve of Dinostratus and Nicomedes. He then proceeds to describe that curve, which is the one now referred to as the quadratrix of Hippias.³

The quadratrix is described in terms of two independent motions of a point as follows. The radius of a circle rotates at a uniform rate from the vertical position AB in Fig. 11.8 to the horizontal position $A\Delta$, while in exactly the same time a horizontal line moves downward at a constant speed from the position $B\Gamma$ to the position $A\Delta$. The point of intersection Z traces the curve BZH , which is the quadratrix. The diameter of the circle is the mean proportional between its circumference and the line AH . Unfortunately, H is the one point on the quadratrix that is not determined, since the two intersecting lines coincide when they both reach $A\Delta$. This point was noted by Pappus, citing an earlier author named Sporos. In order to draw the curve, which is mechanical, you first have to know the ratio of the circumference of a circle to its diameter. But if you knew that, you would already be able to square the circle. One can easily see, however, that since the angle $Z\Delta A$ is proportional to the height of Z , this curve—if it can be drawn!—makes it possible to divide an angle into any number of equal parts.

Pappus also attributed a trisection to Menelaus of Alexandria (70–130 CE). Pappus gave a classification of geometric construction problems in terms of three categories: planar, solid, and [curvi]linear. The first category consisted of constructions that used only straight lines and circles, whereas the second category consisted of those that used conic sections. The last, catch-all category consisted of problems requiring all manner of more elaborate and less regular curves, which were harder to visualize than the first two and presumably required some mechanical device to draw them. We are all familiar with the use of a compass to draw a perfect circle and the procedure for drawing an ellipse by stretching a thread between two

³Hippias should be thankful for Proclus, without whom he would apparently be completely forgotten, as none of the other commentators discuss him, except for a mention in passing by Diogenes Laertius in his discussion of Thales. Allman (1889, pp. 94–95) argued that the Hippias mentioned in connection with the quadratrix is not the Hippias of Elis (ca. 460–ca. 400 BCE) mentioned in the Eudemian summary, and other historians, including the late Wilbur Knorr, have agreed with him, but most do not.

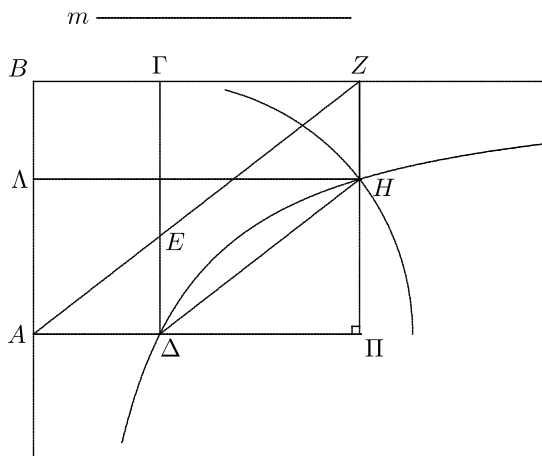


Figure 11.9. Pappus' construction of a *neûsis* using a rectangular hyperbola.

pins at the foci. A device for drawing hyperbolas with given foci is also easy to design.⁴ Somewhat surprisingly, it was not until the nineteenth century that a mechanical device (the Peaucellier linkage, named after Charles-Nicolas Peaucellier, 1832–1914) was invented that draws a theoretically perfect straight line.

The quadratrix described above, however, cannot be drawn with any such instrument; it requires coordinating two independent motions with infinite precision, a thing that is difficult to imagine. Pappus says that some of these more general curves come from locus problems; he goes on to say that geometers regard it as a major defect when a planar problem is solved using conics and other curves.

Based on this classification of problems, the first geometers were unable to solve the above-mentioned problem of [trisecting] the angle, which is by nature a solid problem, through planar methods. For they were not yet familiar with the conic sections; and for that reason they were at a loss. But later they trisected the angle through conics, using the convergence described below.

The word *convergence* (*neûsis*) comes from the verb *neúein*, one of whose meanings is *to incline toward*. In this particular case, it refers to the following construction. We are given a rectangle $AB\Gamma\Delta$ and a prescribed length m . It is required to find a point E on $\Gamma\Delta$ such that when AE is drawn and extended to meet the extension of $B\Gamma$ at a point Z , the line EZ will have length m . The construction is shown in Fig. 11.9, where the circular arc with center at Δ has radius m . The hyperbola is rectangular, with asymptotes BA and BZ , so that $A\Delta \cdot \Gamma\Delta = \Lambda H \cdot ZH$. This equation implies that $\Pi Z : HZ = \Pi A : \Delta A$. Thus the triangles $\Delta\Pi H$ and $A\Pi Z$ are similar, and so AZ is parallel to ΔH , from which it follows that $EZ = \Delta H = m$.

⁴Imagine two spools with meshing gears on axes beneath the plane of the hyperbola, with a continuous thread wound around them in opposite directions and passing up over the table through the two foci. As a point on the thread is pulled, the two interlocked spools will both unwind at the same rate, keeping the difference between the lengths of thread from the given point to the two foci constant. Hence the point will describe a hyperbola.

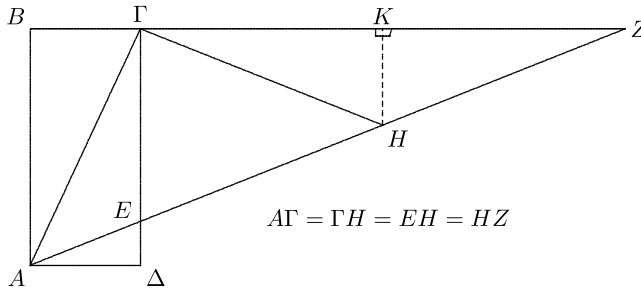


Figure 11.10. Trisection of an arbitrary angle by *neûsis* construction. Because $EH = HZ$ and both HK and $\Gamma\Delta$ are perpendicular to BZ , it follows that $\Gamma K = KZ$. Hence ZKH is congruent to ΓKH , and so $\angle H\Gamma K = \angle HZK = \angle EA\Delta$. But then $\Gamma H = HZ = A\Gamma$, and so $\angle \Gamma A H = \angle \Gamma H A = 2\angle HZ\Gamma$.

With the *neûsis* construction, it becomes a simple matter to trisect an angle, as Pappus pointed out. Given any acute angle, label its vertex A , choose an arbitrary point Γ on one of its sides, and let Δ be the foot of the perpendicular from Γ to the other side of the angle. Complete the rectangle $AB\Gamma\Delta$, and carry out the *neûsis* with $m = 2A\Gamma$. Then let H be the midpoint of ZE , and join ΓH , as shown in Fig. 11.10.

11.3.1. A Mechanical Solution: The Conchoid

Finding the point E in the *neûsis* problem is equivalent to finding the point Z . Either point allows the line AEZ to be drawn. Now one line that each of these points lies on is known. If some other curve that Z must lie on could be drawn, the intersection of that curve with the line $B\Gamma$ would determine Z and hence solve the *neûsis* problem. If we use the condition that the line ZE must be of length $2A\Gamma$, we have a locus-type condition for Z , and it is easy to build a device that will actually draw this locus. What is needed is the T-shaped frame shown in Fig. 11.11, consisting of two pieces of wood or other material meeting at right angles. The horizontal part of the T has a groove along which a peg (shown as a hollow circle in the figure) can slide. The vertical piece has a fixed peg (shown as a solid circle) at

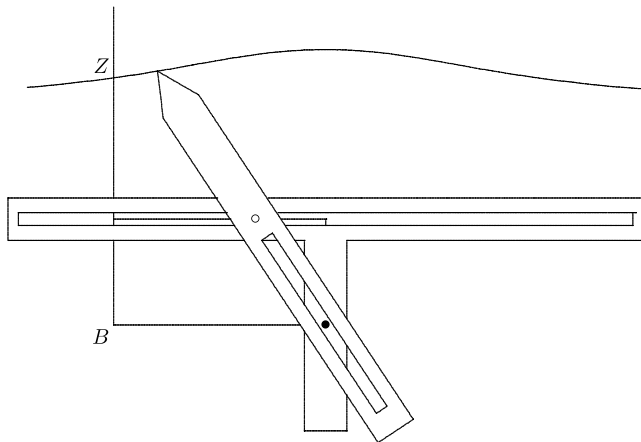


Figure 11.11. A mechanical device for drawing the conchoid of Nicomedes.

distance $A\Delta$ from its top. Onto this frame a third piece is fitted with a fixed peg (the hollow circle) at distance m from its end and a groove between the peg and the other end that fits over the peg on the vertical bar. The frame is then laid down with its horizontal groove over the line $\Gamma\Delta$ and its fixed peg over A . When the moving piece is fitted over the frame so that its peg slides along the horizontal groove over $\Gamma\Delta$ and its groove slides over the peg at A , its endpoint (where a stylus is located to draw the curve) traces the locus on which Z must lie. The point Z lies where that locus meets the extension of $B\Gamma$. In practical terms, such a device can be built, but the rigid pegs must be located at exactly the distance from the ends determined by the rectangle and the fixed distance given in the *neûsis* problem. Thus the device must be modified by moving the pegs to the correct locations for each particular problem. If oxymoron is permitted, we might say that the practical value of this device is mostly theoretical. The locus it draws is the *conchoid of Nicomedes*, mentioned by Pappus and Proclus. (Nothing is known about Nicomedes beyond the facts that he lived during the third century BCE and wrote a treatise on conchoid curves.)

Because of the objections reported by Pappus to the use of methods that were more elaborate than the problems they were intended to solve, the search for planar (ruler-and-compass) solutions to these problems continued for many centuries. It was not until the 1830s that it was proved that no straightedge-and-compass solution exists for any of them. (The proof had no effect on the cranks of the world, of course.) The problems continue to be of interest since that time, and not only to cranks who imagine they have solved them. Felix Klein, a leading German mathematician and educator in the late nineteenth and early twentieth centuries, urged that they be studied as a regular part of the curriculum (Beman and Smith, 1930).

PROBLEMS AND QUESTIONS

Mathematical Problems

- 11.1.** Show why the quantities u , v , w , and h in Fig. 11.6 satisfy the relation $v^2 = h^2 + 2uh - w^2$. (Use the second diagram in the figure, which shows that $v^2 = \overline{RP} \cdot \overline{PT} = (\overline{RC} - \overline{PC}) \cdot (\overline{RC} + \overline{PC}) = \overline{RC}^2 - \overline{PC}^2$. You will also need the relations $(u + h)^2 - \overline{RC}^2 = \overline{OC}^2$ and $\overline{OC}^2 + \overline{PC}^2 = \overline{OP}^2 = u^2 + w^2$. You need to eliminate \overline{OC}^2 using these last two relations.)
- 11.2.** Referring to Fig. 11.5, show that at the intersection of the parabola and hyperbola, where $x = B$ and $y = \Gamma$, we have $A : B :: B : \Gamma$ and $A : B :: \Gamma : E$.
- 11.3.** Explain why the point H in Fig. 11.6 is not determined by the conditions given in the definition of the quadratrix.

Historical Questions

- 11.4.** Why did the center of Greek mathematics shift from the commercial cities in the Ionian Sea to Athens during the fifth century BCE?
- 11.5.** Who were the scholars who came to Athens in fifth and fourth centuries BCE and worked on mathematical problems while they were there?

- 11.6.** Summarize the progress made on each of the three classical problems during the fourth century BCE.

Questions for Reflection

- 11.7.** Try to design a mechanical instrument that will draw a quadratrix for a given circle. (You will need to assume an ideal thread that is perfectly flexible but incapable of being stretched in order to do this.)
- 11.8.** Why is a *neûsis* not a straightedge-and-compass construction?
- 11.9.** Why was it important to Menaechmus' solution of the problem of two mean proportionals that the plane cutting the cone be at right angles to one of its generators?

Athenian Mathematics II: Plato and Aristotle

As we have already mentioned, Plato met the Pythagorean Philolaus in Sicily in 390. He also met the Pythagorean Archytas at Tarentum, where some Pythagoreans had once fled to escape danger at Croton. Plato returned to Athens and founded the Academy in 387 BCE. There he hoped to train the young men¹ for public service and establish good government. At the behest of Archytas and a Syracusan politician named Dion, brother-in-law of the ruler Dionysus I, Plato made several trips to Syracuse (Sicily) between 367 and 361 BCE to act as advisor to Dionysus II. However, there was virtual civil war between Dion and Dionysus, and Plato was arrested and nearly executed. Diogenes Laertius quotes a letter allegedly from Archytas to Dionysus urging that Plato be released. Plato returned to the Academy in 360 and remained there for the last 13 years of his life. He died in 347.

12.1. THE INFLUENCE OF PLATO

Archytas' solution of the problem of two mean proportionals using two half-cylinders intersecting at right angles was mentioned above. In his *Symposium Discourses*, Plutarch claimed that

Plato also lamented that the disciples of Eudoxus, Archytas, and Menaechmus attacked the duplication of a solid by building tools and machinery hoping to get two ratios through the irrational, by which it might be possible to succeed, [saying that by doing so they] immediately ruined and destroyed the good of geometry by turning it back toward the physical and not directing it upward or striving for the eternal and incorporeal images, in which the divinity is eternally divine.

Although the sentiment Plutarch ascribes to Plato is consistent with the ideals expressed in the *Republic*, Eutocius reports one such mechanical construction as being due to Plato

¹In his writing, especially *The Republic*, Plato argues for equal participation by women in government. There is no record of any women students at his Academy, however. His principles were far in advance of what the Athenians would tolerate in practice.

himself. From his upbringing as a member of the Athenian elite and from the influence of Socrates, Plato had a strong practical streak, concerned with life as it is actually lived.² Platonic idealism in the purely philosophical sense does not involve idealism in the sense of unrealistic striving for perfection.

It may have been Archytas and Philolaus who aroused Plato's interest in mathematics, an interest that continued for the rest of his life. Mathematics played an important role in the curriculum of his Academy and in the research conducted there. Lasserre (1964, p. 17) believes that the most important mathematical work at the Academy began with the arrival of Theaetetus in Athens around 375 and ended with Eudoxus' departure for Cnidus around 350.

The principle that knowledge can involve only eternal, unchanging entities led Plato to some statements that sound paradoxical. For example, in Book 7 of the *Republic* he writes:

Thus we must make use of techniques such as geometry when we take up astronomy and ignore what is in the sky if we really intend to create something intrinsically useful and practical in the soul.

If Plato's mathematical concerns seem to be largely geometrical, that is probably because he became acquainted with mathematics at the time when the challenges discussed above were still current topics. (Recall the quotation from the *Republic* in Chapter 8, where he laments the lack of public support for research into solid geometry.) There is a long-standing legend that Plato's Academy bore the following sign above its entrance³:

ΑΓΕΩΜΕΤΡΗΤΟΣ ΜΗΔΕΙΣ ΕΙΣΙΤΩ

(*AGEŌMETRĒTOS MĒDEIS EISITŌ*, that is, "Let no one unskilled in geometry enter.") If Plato really was more concerned with geometry than with arithmetic, there is an obvious explanation for his preference: The imperfections of the real world come more from geometry than arithmetic. For example, it is sometimes asserted that there are no examples of exact equality in the real world. But in fact, as was pointed out in Chapter 1, there are many. Those who make the assertion always have in mind continuous magnitudes, such as lengths or weights, in other words, geometrical concepts. Where arithmetic is concerned, exact equality is easy to achieve, as shown by the example of equal bank accounts in Chapter 1. But Plato's love for geometry should not be overemphasized. In his ideal curriculum, described in the *Republic*, arithmetic is still regarded as the primary subject.

²In the famous allegory of the cave in Book 7 of the *Republic*, Plato depicts the unphilosophical person as living in a cave with feet in chains, seeing only flickering shadows on the wall of the cave, while the philosopher is the person who has stepped out of the cave into the bright sunshine and wishes to communicate that reality to the people back in the cave. While he encouraged his followers to "think outside the cave," his trips to Syracuse show that he understood the need to make philosophy work inside the cave, where everyday life was going on.

³These words are the earliest version of the legend, which Fowler (1998, pp. 200–201) found could not be traced back earlier than a scholium attributed to the fourth-century orator Sopatros. The commonest source cited for this legend is the twelfth-century Byzantine Johannes Tzetzes, in whose *Chiliades*, VIII, 975, one finds *Μηδεις ἀγεωμέτρητος εἰσῆτω μου τὴν στέγην*. "Let no one unskilled in geometry enter my house."

12.2. EUDOXAN GEOMETRY

We recall the difficulty occasioned for the theory of proportion by the discovery of incommensurables, as illustrated by Fig. 4 of Chapter 10. The solution to this difficulty was provided by Eudoxus of Cnidus (ca. 407–354 BCE), whom Diogenes Laertius describes as “astronomer, geometer, physician, and lawgiver.” He learned geometry from Archytas and philosophy from Plato. Diogenes Laertius cites another commentator, named Sotion, who said that Eudoxus spent two months in Athens and attended lectures by Plato. Because of his poverty, he could not afford to live in Athens proper. He lived at the waterfront, known as the Piraeus, supported by a physician named Theomedus, and walked 11 km from there into Athens. Then, with a subsidy from friends, he went to Egypt and other places and finally returned, “crammed full of knowledge,” to Athens, “some say, just to annoy Plato for snubbing him earlier.” Plato was not in Eudoxus’ league as a mathematician; and if Eudoxus felt that Plato had patronized him in his earlier visit, perhaps because Plato and his other students were wealthy and Eudoxus was poor, his desire to return and get his own students back from Plato is quite understandable. He must have made an impression on Plato on his second visit. In his essay *On Socrates’ Daemon*, Plutarch reports that when the Delians consulted Plato about doubling the cube, in addition to advising them to study geometry, he told them that the problem had already been solved by Eudoxus of Cnidus and Helicon of Cyzicus. If true, this story suggests that the Delians appealed to Plato after Eudoxus had left for Cnidus, around 350. In Cnidus, Eudoxus made many astronomical observations that were cited by the astronomer Hipparchus (ca. 190–ca. 120 BCE), and one set of his astronomical observations has been preserved. Although the evidence is not conclusive, it seems that while he was in Athens, he contributed two vital pieces to the mosaic that is Euclid’s *Elements*.

12.2.1. The Eudoxan Definition of Proportion

The first piece of the *Elements* probably contributed by Eudoxus was the solution of the problem of incommensurables. This solution is attributed to him on the basis of two facts: (1) Proclus’ comment that Euclid “arranged many of the theorems of Eudoxus”; (2) an anonymous scholium (commentary) on Euclid’s Book 5, which asserts that the book is the creation “of a certain Eudoxus, [the student] of the teacher Plato” (Allman, 1889, p. 132).

The main principle is very simple: Suppose that D and S are, respectively, the diagonal and side of a square or pentagon. Even though there are no integers m and n such that $mD = nS$, so that the ratio $D : S$ cannot be defined as $n : m$ for any integers, it remains true that for every pair of integers m and n there is a trichotomy: Either $mD < nS$ or $mD = nS$ or $mD > nS$. That fact makes it possible at least to define what is meant by saying that the ratio of D to S is the same for all similar polygons. We define the proportion $D_1 : S_1 :: D_2 : S_2$ for two different squares to mean that, for any positive integers m and n , whatever relation holds between mD_1 and nS_1 also holds between mD_2 and nS_2 . That is, if $mD_1 > nS_1$, then $mD_2 > nS_2$, and similarly for the opposite inequality or equality.

As defined by Euclid at the beginning of Book 5, “A relation that two magnitudes of the same kind have due to their sizes is a *ratio*.” As a definition, this statement is somewhat lacking, but we may paraphrase it as follows: “the relative size of one magnitude in terms of a second magnitude of the same kind is the *ratio* of the first to the second.” We think of size as resulting from measurement and relative size as the result of *dividing* one measurement

by another, but Euclid keeps silent on both of these points. Then, “Two magnitudes are said to *have a ratio to each other* if they are capable of exceeding each other when multiplied.” That is, some (positive integer) multiple of each is larger than the other. Thus, the periphery of a circle and its diameter have a ratio, but the periphery of a circle and the disk it encloses do not. Although this definition of ratio would be hard to use, fortunately there is no need to use it. What is needed is equality of ratios, that is, proportion. That definition follows from the trichotomy just mentioned. Here is the definition given in Book 5 of Euclid, with the material in brackets added from the discussion just given to clarify the meaning:

Magnitudes are said to be in the same ratio, the first to the second [$D_1 : S_1$] and the third to the fourth [$D_2 : S_2$], when, if any equimultiples whatever be taken of the first and third [mD_1 and mD_2] and any equimultiples whatever of the second and fourth [nS_1 and nS_2], the former equimultiples alike exceed, are alike equal to, or are alike less than the latter equimultiples taken in corresponding order [that is, $mD_1 > nS_1$ and $mD_2 > nS_2$, or $mD_1 = nS_1$ and $mD_2 = nS_2$, or $mD_1 < nS_1$ and $mD_2 < nS_2$].

Let us now revisit our conjectured early proof of Euclid’s Proposition 1 of Book 6 of the *Elements* from Chapter 10, a proof that holds only in the commensurable case. How much change is required to make this proof cover the incommensurable case? Very little, as it turns out. Where we have assumed that $3BC = 2CD$, it is only necessary to consider the cases $3BC > 2CD$ and $3BC < 2CD$ and show with the same figure that $3ABC > 2ACD$ and $3ABC < 2ACD$, respectively, and that is done by using the trivial corollary of Proposition 38 of Book 1: *If two triangles have equal altitudes and unequal bases, the one with the larger base is larger*. Eudoxus has not only shown how proportion can be defined so as to apply to incommensurables, he has done so in a way that fits together seamlessly with earlier proofs that apply only to the commensurable case. If only the fixes for bugs in modern computer programs were so simple and effective!

12.2.2. The Method of Exhaustion

Eudoxus’ second contribution is of equal importance with the first; it is the proof technique known as the *method of exhaustion*. This method is used by both Euclid and Archimedes to establish theorems about areas and solids bounded by curved lines and surfaces. As in the case of the definition of proportion for incommensurable magnitudes, the evidence that Eudoxus deserves the credit for this technique is not conclusive. In his commentary on Aristotle’s *Physics*, Simplicius credits the Sophist Antiphon with inscribing a polygon in a circle, then repeatedly doubling the number of sides in order to square the circle. However, the perfected method seems to belong to Eudoxus. Archimedes says in the cover letter accompanying his treatise on the sphere and cylinder that it was Eudoxus who proved that a pyramid is one-third of a prism on the same base with the same altitude and that a cone is one-third of the cylinder on the same base with the same altitude. What Archimedes meant by proof we know: He meant proof that meets Euclidean standards. Such a proof can be achieved for the cone only by the method of exhaustion. Like the definition of proportion, the basis of the method of exhaustion is a simple observation: When the number of sides of a polygon inscribed in a circle is doubled, the excess of the circle over the polygon is reduced by more than half, as one can easily see from Fig. 12.1. This observation works together with the theorem that *if two magnitudes have a ratio and more than half of the larger is removed, then more than half of what remains is removed*, and this process continues, then

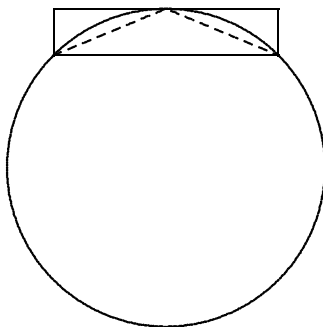


Figure 12.1. The basis of the method of exhaustion.

at some point what remains will be less than the smaller of the original two magnitudes (*Elements*, Book 10, Proposition 1). This principle is usually called *Archimedes' principle* because of the frequent use he made of it. The phrase *if two magnitudes have a ratio* is critical, because Euclid's proof of the principle depends on converting the problem to a problem about integers.

To be specific, if $a > b$, since $nb > a$ for some positive integer n , it is only a matter of showing that a finite sequence $a = a_1, a_2, \dots$ in which each term is less than half of the preceding will eventually reach a term a_k such that na_k is less than a_1 . Since $nb > a = a_1$, it follows that $a_k < b$. Since $m/2^m < 1$ for $m > 1$, we see that in fact k will be less than or equal to n .

The definition of ratio and proportion allowed Eudoxus/Euclid to establish all the standard facts about the theory of proportion, including the important fact that similar polygons are proportional to the squares on their sides (*Elements*, Book 6, Propositions 19 and 20). Once that result is achieved, the method of exhaustion makes it possible to establish rigorously that similar curvilinear regions are proportional to the squares on similarly situated chords. In particular, it made it possible to prove the fundamental fact that was being used by Hippocrates much earlier: Circles are proportional to the squares on their diameters. This fact is now stated as Proposition 2 of Book 12 of the *Elements*, and the proof given by Euclid is illustrated in Fig. 12.2.

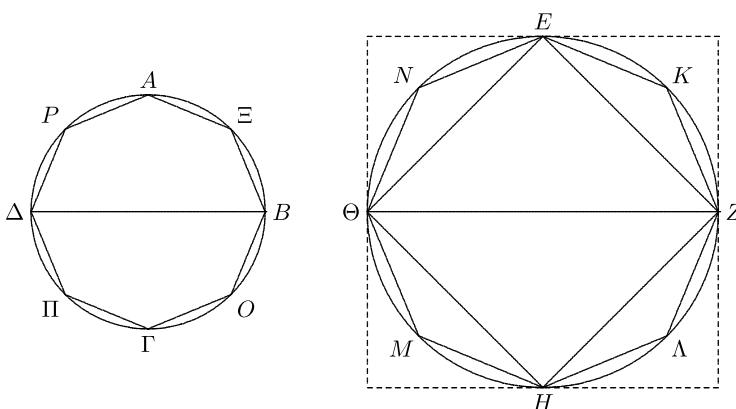


Figure 12.2. Proof that circles are proportional to the squares on their diameters.

Let $AB\Gamma\Delta$ and $EZH\Theta$ be two circles with diameters $B\Delta$ and ΘZ , and suppose that the circles are *not* proportional to the squares on their diameters. Let the ratio $B\Delta^2 : \Theta Z^2$ be the same as $AB\Gamma\Delta : \Sigma$, where Σ is an area larger or smaller than $EZH\Theta$. Suppose first that Σ is smaller than the circle $EZH\Theta$. Draw the square $EZH\Theta$ inscribed in the circle $EZH\Theta$. Since this square is half of the circumscribed square with sides perpendicular and parallel to the diameter ΘZ , and the circle is smaller than the circumscribed square, the inscribed square is more than half of the circle. Now bisect each of the arcs EZ , ZH , $H\Theta$, and ΘE at points K , Λ , M , and N , and join the polygon $EKZ\Lambda H\Theta NE$. As shown above, doing so produces a larger polygon, and the excess of the circle over this polygon is less than half of its excess over the inscribed square. If this process is continued enough times, the excess of the circle over the polygon will eventually be less than its excess over Σ , and therefore the polygon will be larger than Σ . For definiteness, Euclid assumes that this polygon is the one reached at the first doubling: $EKZ\Lambda H\Theta NE$. In the first circle $AB\Gamma\Delta$, inscribe a polygon $A\Xi BO\Gamma\Pi\Delta P$ similar to $EKZ\Lambda H\Theta NE$. Now the square on $B\Delta$ is to the square on $Z\Theta$ as $A\Xi BO\Gamma\Pi\Delta P$ is to $EKZ\Lambda H\Theta NE$. But also the square on $B\Delta$ is to the square on $Z\Theta$ as the circle $AB\Gamma\Delta$ is to Σ . It follows that $A\Xi BO\Gamma\Pi\Delta P$ is to $EKZ\Lambda H\Theta NE$ as the circle $AB\Gamma\Delta$ is to Σ . Since the circle $AB\Gamma\Delta$ is larger than $A\Xi BO\Gamma\Pi\Delta P$, it follows that Σ must be larger than $EKZ\Lambda H\Theta NE$. But by construction, it is smaller, which is impossible. A similar argument shows that it is impossible for Σ to be larger than $EZH\Theta$.

12.2.3. Ratios in Greek Geometry

Ratios as defined by Euclid are always between two magnitudes of the same type. He never considered what we call density, for example, which is the ratio of a mass to a volume. Being always between two magnitudes of the same type, ratios are “dimensionless” in our terms and could be used as numbers, if only they could be added and multiplied. The Greeks, however, did not think of these operations on ratios as being the same thing they could do with numbers. In terms of adding, Euclid does say (Book 6, Proposition 24) that if two proportions have the same second and fourth terms, then their first terms and third terms can be added (first to first and third to third), that is, if $a : b :: c : d$ and $e : b :: f : d$, then $(a + e) : b :: (c + f) : d$. But he did not think of the second and fourth terms in a proportion as denominators, and this was not, as we see it, merely adding fractions with a common denominator. For multiplication of ratios, Euclid gives three separate definitions. In Book 5, Definition 9, he defines the duplicate (which we would call the square) of the ratio $a : b$ to be the ratio $a : c$ if b is the mean proportional between a and c , that is, $a : b :: b : c$. Similarly, when there are four terms in proportion, as in the problem of two mean proportionals, so that $a : b :: b : c :: c : d$, he calls the ratio $a : d$ the triplicate of $a : b$. We would call it the cube of this ratio. Not until Book 6, Definition 5 is there any kind of general definition of the product of two ratios. Even that definition is not in all manuscripts and may be a later interpolation. It goes as follows: *A ratio is said to be the composite of two ratios when the sizes in the two ratios produce something when multiplied by themselves.*⁴ This rather vague definition is made still harder to grasp by the fact that the word for *composite* (*sygkeímena*) is simply a general word for *combined*. It means literally *lying together* and is the same word used when two lines are placed end to end to form a longer line. In that context it

⁴I am aware that the word “in” here is not a literal translation, since the Greek has the genitive case—the sizes of the two ratios. But I take *of* here to mean *belonging to*, which is one of the meanings of the genitive case.

corresponds to addition, whereas in the present one it corresponds to multiplication. It can be understood only by seeing the way that Euclid operates with it. Given four lines a , b , c , and d , to form the composite ratio $a : b.c : d$, Euclid first takes any two lines⁵ x and y such that $a : b :: x : y$. He then finds a line z such that $y : z :: c : d$ and defines the composite ratio $a : b.c : d$ to be the ratio $x : z$.

There is some arbitrariness in this procedure, since x could be any line. A modern mathematician looking at this proof would note that Euclid could have shortened the labor by taking $x = a$ and $y = b$, then constructing z to have the same ratio to y that d has to c . The same mathematician would add that Euclid ought to have shown that the final ratio is the same independently of the choice of x , which he did not do. But one must remember that the scholarly community around Euclid was much more intimate than in today's world; he did not have to write a "self-contained" book. In the present instance a glance at Euclid's *Data* shows that he knew what he was doing. The first proposition in that book says that "if two magnitudes [of the same kind] A and B are given, then their ratio is given." In modern language, any quantity can be replaced by an equal quantity in a ratio without changing the ratio. The proof is that if $A = \Gamma$ and $B = \Delta$, then $A : \Gamma :: B : \Delta$, and hence by Proposition 16 of Book 5 of the *Elements*, $A : B :: \Gamma : \Delta$. The second proposition of the *Data* draws the corollary that if a given magnitude has a given ratio to a second magnitude, then the second magnitude is also given. That is, if two quantities have the same ratio to a given quantity, then they are equal. From these principles, Euclid could see that the final ratio $x : z$ is what mathematicians now call "well-defined," that is, independent of the choice of x .⁶ The first use made of this process is in Proposition 23 of Book 6, which asserts that equiangular parallelograms are in the compound ratio of their (corresponding) sides.

With the departure of Eudoxus for Cnidus, we can bring to a close our discussion of Plato's influence on mathematics. If relations between Plato and Eudoxus were less than intimate, as Diogenes Laertius implies, Eudoxus may have drawn off some of Plato's students whose interests were more scientific (in modern terms) and less philosophical. It is likely that even Plato realized that his attempt to explain the universe by means of eternal ideal forms, for the understanding of which mathematics was a useful training tool, would not work after all. His late dialogue *Parmenides* gives evidence of a serious rethinking of this doctrine.

12.3. ARISTOTLE

Plato died in 347 BCE, and his place as the preeminent scholar of Athens was taken a decade after his death by his former pupil Aristotle (384–322 BCE). Aristotle became a student at the Academy at the age of 18 and remained there for 20 years. After the death of Plato he left Athens, traveled, got married, and in 343 became tutor to the future Macedonian King Alexander (the Great), who was 13 years old when Aristotle began to teach him and 16 when he became king on the death of his father. In 335 Aristotle set up his own school,

⁵ Actually, m and n need not be lines as long as they "have a ratio," that is, are geometric objects of the same kind. The exact nature of x , y , and z is not important, since in applications only the integers by which they are multiplied play any role in the argument.

⁶ A good exposition of the purpose of Euclid's *Data* and its relation to the *Elements* was given by Il'ina (2002), elaborating a thesis of I. G. Bashmakova.

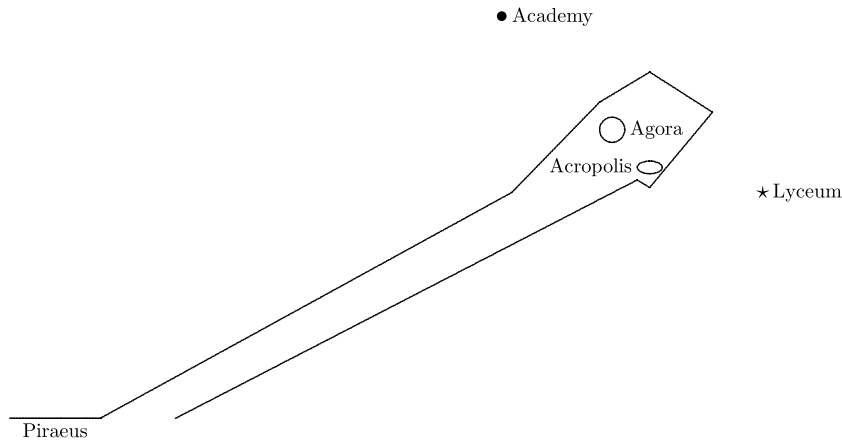


Figure 12.3. Athens in the fourth century BCE: the waterfront (Piraeus), Academy, and Lyceum.

located in the Lyceum, over the hill from the Academy (Fig. 3).⁷ For the next 12 years he lived and wrote there, producing an enormous volume of speculation on a wide variety of subjects, scientific, literary, and philosophical. In 322 Alexander died, and the Athenians he had conquered turned against his friends. Unlike Socrates, Aristotle felt no obligation to be a martyr to the laws of the *polis*. He fled to escape the persecution, but died the following year. Aristotle's writing style resembles very much that of a modern scholar, except for the absence of footnotes. Like Plato, in mathematics he seems more like a well-informed generalist than a specialist.

The drive toward the logical organization of science reached its full extent in the treatises of Aristotle. He analyzed reason itself and gave a rigorous discussion of formal inference and the validity of various kinds of arguments in his treatise *Prior Analytics*, which was written near the end of his time at the Academy, around 350 BCE. One can almost picture debates at the Academy, with the mathematicians providing examples of their reasoning, which the logician Aristotle examined and criticized in order to distill his rules for making inferences. In this treatise Aristotle discusses subjects, predicates, and syllogisms connecting the two, occasionally giving a glimpse of some mathematics that may indicate what the mathematicians were doing at the time.

In Book 1 of the *Prior Analytics*, Aristotle describes how to organize the study of a subject, looking for all the attributes and subjects of both of the terms that are to appear in a syllogism. The subject–attribute relation is mirrored in modern thought by the notion of elements belonging to a set. The element is the subject, and the set it belongs to is defined by attributes that can be predicated of all of its elements and no others. Just as sets can be elements of other sets, Aristotle said that the same object can be both a subject and a predicate. He thought, however, that there were some absolute subjects (individual people, for example) that were not predicates of anything and some absolute predicates (what we

⁷The names of these two institutions have become a basic part of our intellectual world, masking their origins. Both are named for their geographical location in Athens. The Academy was a wooded area named in honor of Akademos, who, according to legend, saved Athens from the wrath of Castor and Pollux by telling them where the Athenian king Theseus had hidden their sister Helen. The Lyceum was located near the temple of Apollo Lykeios (“Apollo of the Wolves”).

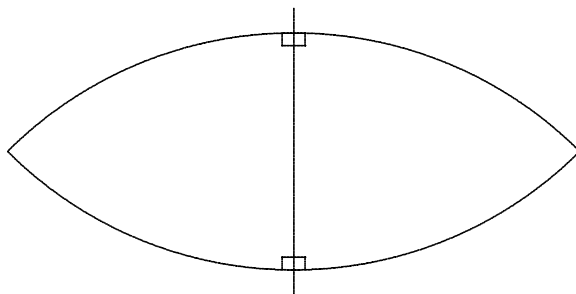


Figure 12.4. How do we exclude the possibility that two lines perpendicular to the same line may intersect each other?

call abstractions, such as beauty) that were never the subject of any proposition.⁸ Aristotle says that the postulates appropriate to each subject must come from experience. If we are thorough enough in stating all the attributes of the fundamental terms in a subject, it will be possible to prove certain things and state clearly what must be assumed.

In Book 2 he discusses ways in which reasoning can go wrong, including the familiar fallacy of “begging the question” by assuming what is to be proved. In this context he offers as an example the people who claim to construct parallel lines. According to him, they are begging the question, starting from premises that cannot be proved without the assumption that parallel lines exist. We may infer that there were around him people who did claim to show how to construct parallel lines, but that he was not convinced. It seems obvious that two lines perpendicular to the same line are parallel, but surely that fact, so obvious to us, would also be obvious to Aristotle. Therefore, he must have looked beyond the obvious and realized that the existence of parallel lines does *not* follow from the immediate properties of lines, circles, and angles. Only when this realization dawns is it possible to see the fallacy in what appears to be common sense. Common sense—that is, human intuition—suggests what can be proved: If two perpendiculars to the same line meet on one side of the line, then they must meet on the other side also, as in Fig. 12.4. Indeed, Ptolemy did prove this, according to Proclus. But Ptolemy then concluded that two lines perpendicular to the same line cannot meet at all. “But,” Aristotle would have objected, “you have not proved that two lines cannot meet in two different points.” And he would have been right: The assumptions that two lines can meet in only one point and that the two sides of a line are different regions (not connected to each other) are equivalent to assuming that parallel lines exist.

Euclid deals with this issue in the *Elements* by stating as the last of his assumptions that “two straight lines do not enclose an area.” Oddly, however, he seems unaware of the need for this assumption when proving the main lemma (Book 1, Proposition 16) needed to prove the existence of parallel lines.⁹ This proposition asserts that an exterior angle of a triangle is larger than either of the opposite interior angles. Euclid’s proof is based on

⁸In modern set theory it is necessary to assume that one cannot form an infinite chain of sets a, b, c, \dots such that $b \in a, c \in b, \dots$. That is, at some finite stage in such a chain of element relations, there is an “atom” that has no elements, what is called the empty set.

⁹In standard editions of Euclid, there are 14 assumptions, but three of them, concerned with adding equals to equals, doubling equals, and halving equals, are not found in some manuscripts. Gray (1989, p. 46) notes that the fourteenth assumption may be an interpolation by the Muslim mathematician al-Nayrizi, (ca. 875–ca. 940) the result of speculation on the foundations of geometry. That would explain its absence from the proof of Proposition 16.

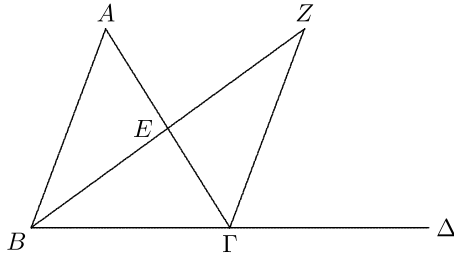


Figure 12.5. The exterior angle theorem (*Elements*, I, 16).

Fig. 12.5, in which a triangle $AB\Gamma$ is given with side $B\Gamma$ extended to Δ , forming the exterior angle $A\Gamma\Delta$. He wishes to prove that this angle is larger than the angle at A . To do so, he bisects $A\Gamma$ at E , draws AE , and extends it to Z so that $EZ = AE$. When $Z\Gamma$ is joined, it is seen that the triangles ABE and ΓZE are congruent by the side–angle–side criterion. It follows that the angle at A equals $\angle E\Gamma Z$, which is smaller than $\angle E\Gamma\Delta$, being only a part of it.

In the proof, Euclid assumes that the points E and Z are on the same side of line $B\Gamma$. But that is obvious only for triangles small enough to see. It needs to be proved. To be sure, Euclid could have proved it by arguing that if E and Z were on opposite sides of $B\Gamma$, then EZ would have to intersect either $B\Gamma$ or its extension in some point H , and then the line BH passing through Γ and the line BEH would enclose an area. But he did not do that. In fact, the only place where Euclid invokes the assumption that two lines cannot enclose an area is in the proof of the side–angle–side criterion for congruence (Book 1, Proposition 4).¹⁰

Granting that Aristotle was right about this point, we still must wonder why he considered the existence of parallel lines to be in need of proof. Why would he have doubts about something that is so clear on an intuitive level? One possible reason is that parallelism involves the infinite: Parallel lines will *never* meet, no matter how far they are extended. If geometry is interpreted physically (say, by regarding a straight line as the path of a light ray), we really have no assurance whatever that parallel lines exist—how could anyone assert with confidence what will happen if two apparently parallel lines are extended to a length of hundreds of light years?

As Aristotle's discussion of begging the question continues, further evidence comes to light that this matter of parallel lines was being debated around 350, and proofs of the existence of parallel lines (Book 1, Proposition 27 of the *Elements*) were being proposed, based on the exterior-angle principle. In pointing out that different false assumptions may lead to the same wrong conclusion, Aristotle notes in particular that the nonexistence of parallel lines would follow if an internal angle of a triangle could be greater than an external angle (not adjacent to it), and also if the angles of a triangle added to more than two right angles.¹¹ One is almost tempted to say that the mathematicians who analyzed the matter in this way foresaw the non-Euclidean geometry of Riemann, but of course that could not be. Those mathematicians were examining what must be assumed in order to get parallel lines into their geometry. They were not exploring a geometry without parallel lines.

¹⁰This proof also uses some terms and some hidden assumptions that are visually obvious but which mathematicians nowadays insist on making explicit.

¹¹Field and Gray (1987, p. 64) note that this point has been made by many authors since Aristotle.

It is precisely in the matter of the parallel postulate—equivalent to the angle sum of a triangle being two right angles—that we see the extent to which geometry was still partly an intuitive science, not merely a matter of verbal deduction from premises. Nowadays we take it for granted that formal systems begin with undefined terms and axioms and proceed to deduce theorems. As far as mathematics is concerned, the undefined terms have no interpretation. But for centuries they did have interpretation, and Aristotle shows what it was in a passage from his *Physics* (Bekker, 200a).

There are closely similar inevitable paths both in mathematics and in the natural world. For if the three sides [of a triangle] are straight lines, then [the sum of the angles of] the triangle is two right [angles]; and if the latter holds, so does the former. But if the latter does not hold, then the lines are not straight.

This passage gives a hint that Aristotle knew about the angles of spherical triangles and knew that they added up to more than two right angles. Thus, although *we* can interpret the word *line* to mean a great circle on a sphere, Aristotle could not, since a line was not merely an undefined term for him.

PROBLEMS AND QUESTIONS

Mathematical Problems

- 12.1.** How would you establish that two triangles with equal altitudes and equal bases are equal (“in area,” as we would say, although Euclid would not)?
- 12.2.** Rephrase Hippocrates’ quadrature of the lune shown in Fig. 1 of Chapter 11 in terms of proportions between figures and the squares on parts of their boundaries. (In other words “de-algebraize” the argument given in the text of the preceding chapter.)
- 12.3.** Consider two rectangles R with sides a and b and S with sides c and d . Prove that the ratio $R : S$ is the composite of the two ratios $a : c$ and $b : d$. [*Hint:* Assume given three quantities x , y , and z such that $a : c :: x : y$ and $b : d :: y : z$. You need to prove that $R : S :: x : z$. To that end, let m and n be any positive integers such that $mx > nz$. You now need to prove that $mR > nS$. In order to do that, choose an integer p so large that $p(mx - nz) > y$. Then there is an integer q such that $pmx > qy > pnz$. (Why?) It follows from the proportions that $pma > qc$ and $qb > pnd$. Hence the rectangle with sides pma and qb is larger than the rectangle with sides qc and pnd . Show how to conclude from that result that $mR > nS$.]

Historical Questions

- 12.4.** In what important ways, in addition to the logical arrangement of materials and the insistence on strict proof, does Greek mathematics differ from the mathematics of Mesopotamia and Egypt?
- 12.5.** Why did Plato regard the study of mathematics as important for those who were to be the guardians of his ideal state?

- 12.6.** Why is the problem of squaring the circle much more difficult than the problem of doubling the cube or trisecting the angle?

Questions for Reflection

- 12.7.** Why was the problem of incommensurables a genuine difficulty that needed to be overcome, while the paradoxes of Zeno and the classical construction problems were not?
- 12.8.** It appears that the Greeks overlooked a simple point that might have led them to break out of the confining circle of Euclidean methods. If only they had realized that composite ratios represent multiplication, they would have been freed from the need for dimensional consistency, since their ratios were dimensionless. They could, for example, multiply any number of ratios, whereas interpreting the product of two lines as a rectangle precluded the possibility of any geometric interpretation of product containing more than three factors. Could they have developed analytic geometry if they had made this realization? What else would they have needed?
- 12.9.** Granting that if two lines perpendicular to the same transversal line meet on one side of that line, reflection about the midpoint of the interval between the two points where the lines meet the transversal shows that they must also meet on the other side. How do you know that these two points of intersection are not the same point? What other assumption must you introduce in order to establish that they are different?

Euclid of Alexandria

In many ways, the third century BCE looks like the high-water mark of Greek geometry. This century saw the creation of sublime mathematics in the treatises of Euclid, Archimedes, and Apollonius. It is very tempting to regard Greek geometry as essentially finished after Apollonius, to see everything that came before as leading up to these creations and everything that came after as “polishing up.” And indeed, although there were some bright spots afterward and some interesting innovations, none had the scope or the profundity of the work done by these three geometers.

The first of the three major figures from this period is Euclid, who is world famous for his *Elements*, which we have in essence already discussed. This work is so famous, and it dominated all teaching in geometry throughout much of the world for so long that the man and his work have essentially merged. For centuries, people did not say that they were studying geometry, but instead that they were studying Euclid. This one work has eclipsed both Euclid’s other books and his biography. He did write other books, two of which still exist, named the *Data* and *Optics*. Other works are ascribed to him by Pappus, including *Phaenomena*, *Loci*, *Conics*, and *Porisms*. Pappus quotes theorems from some of these works.

Euclid is defined for us as the author of the *Elements*. Apart from his writings, we know only that he worked at Alexandria in Egypt after the establishment of the Library there. He was traditionally thought to have flourished about 300 BCE, but, as mentioned in Chapter 8, Alexander Jones argued in favor of a later date. The quotations from ancient authors supporting the traditional date are of doubtful authenticity, and they may be interpolations by later editors. In a possibly spurious passage in Book 7 of his *Synagōgē* (*Collection*), Pappus gives a brief description of Euclid as the most modest of men, a man who was precise but not boastful, like (he implies) Apollonius, whom he disparages.

13.1. THE ELEMENTS

The earliest existing manuscripts of the *Elements* date to the ninth century CE, nearly 1200 years after the book was originally written. These manuscripts have passed by many editors, and some passages seem to have been added by hands other than Euclid’s, especially Theon of Alexandria. Theon was probably not interested in preserving an ancient artifact unchanged; he was more likely trying to produce a good, usable treatise on geometry. Some



The Oxyrhynchus fragment. Courtesy of the Penn Museum, image #142655.

manuscripts have 15 books, but the last two have since been declared spurious by the experts, so that the currently standard edition has 13 books, the last of which looks suspiciously less formal than the first 12, leading some to doubt that Euclid wrote it. For the few bits of evidence that exist concerning the original text, we once again have the dry Egyptian climate to thank, which preserved a few fragments of papyrus. The best example comes from an 1896 excavation at Oxyrhynchus, about 100 km upstream from Cairo, which turned up the fragment of Book 2, Proposition 5, by good fortune, a key proposition from that book (see photo). As mentioned, there are also a few ostraca containing propositions from the *Elements*.

Leaving aside the question of which parts were actually written by Euclid, we shall give a summary of the contents, which we have seen coming together in the work of the fifth and fourth-century mathematicians.

13.1.1. Book 1

The contents of the first book of the *Elements* are covered in the standard geometry courses given in high schools. This material involves the elementary geometric constructions of copying angles and line segments, drawing squares, and the like and the basic properties of parallelograms, culminating in the Pythagorean theorem (Proposition 47) and its converse (Proposition 48). In addition, these properties are applied to the problem of transformation of area, leading to the construction of a parallelogram with a given base angle and having an area equal to that of any given polygon (Proposition 45). There the matter rests until the end of Book 2, where it is shown (Proposition 14) how to construct a square equal to any given polygon.

13.1.2. Book 2

The second book contains geometric constructions needed to solve problems that may involve quadratic incommensurables without resorting to the Eudoxan theory of proportion. For example, a fundamental result is Proposition 5: *If a straight line is cut into equal and unequal segments, the rectangle contained by the unequal segments of the whole together with the square on the straight line between the points of the section is equal to the square*

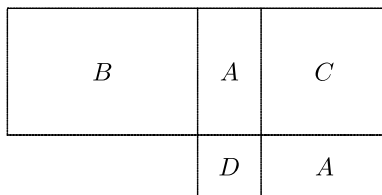


Figure 13.1. Book 2, Proposition 5: Expression of a rectangle as the difference of two squares.

on the half. This proposition is easily seen using Fig. 13.1, in terms of which it asserts that $(A + B) + D = 2A + C + D$; that is, $B = A + C$. It is used as a “polarization” technique, in the form $A + B = (2A + C + D) - D$, where $A + B$ is a rectangle, $2A + C + D$ is the square on the average of its length and width, and D is the square on half of the difference between the length and width.

Here for the first time, the polarization identity and the Pythagorean theorem are combined to express a rectangle as the difference of two squares which in turn is the square on a known line, thus implicitly showing how to transform a rectangle into a square. As we have said, the key ingredients were known in ancient Mesopotamia, but not combined in this way.

Geometric Algebra. We have already mentioned the once-popular interpretation of Book 2 as geometric expressions of what we regard as algebraic identities. It is understandable that people were tempted to think of these propositions in this way. The proposition just stated says $uv = \left(\frac{u+v}{2}\right)^2 - \left(\frac{u-v}{2}\right)^2$. That fact, interpreted numerically, occurs as a fundamental tool in the cuneiform tablets. For if the unequal segments of the line are regarded as two unknown quantities, then half of the original line segment is their average, and the straight line between the points (that is, the segment between the midpoint of the whole segment and the point dividing the whole segment into unequal parts) is what we called earlier the semidifference. Thus, this proposition says that the square of the average equals the product plus the square of the semidifference. Recall that that result was fundamental for solving the important problems of finding two numbers, given their sum and product or their difference and product. On the other hand, in most cases, even in the cuneiform tablets, those two numbers were still the length and width of a rectangle. Thus, who can say whether this proposition was thought of separately from plane geometry? In Euclid, most of the geometric constructions that depend on this proposition (the problems of application with defect and excess) do not appear until Book 6. These application problems could have been solved in Book 2 in the case when the excess or defect is a square. Instead, these special cases were passed over and the more general results, which depend on the Eudoxan theory of proportion to cover the incommensurable cases, were included in Book 6.

Book 2 also contains the construction of what came to be known as the *Section*, that is, the division of a line in mean and extreme ratio so that the whole is to one part as that part is to the other. But Euclid is not ready to prove that version yet, since he doesn’t have the theory of proportion. Instead, he gives what must have been the original form of this proposition (Proposition 11): *Cut a line so that the rectangle on the whole and one of the parts equals the square on the other part.* One way of doing it is shown in Fig. 13.2.

After it is established that four lines are proportional when the rectangle on the means equals the rectangle on the extremes (Proposition 16, Book 6), it becomes possible to convert this construction into the construction of the *Section* (Proposition 30, Book 6).

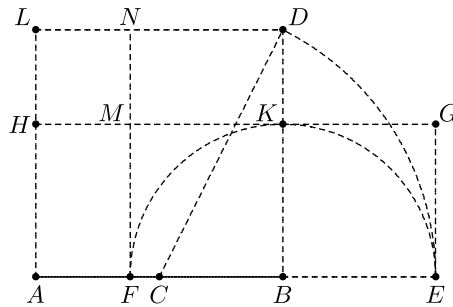


Figure 13.2. Line AB is given with midpoint C , and $ABDL$ is a square. The circular arc \widehat{DE} is centered at C and the semicircle \widehat{FE} is centered at B . Then $AE : AB :: AB : BF :: BF : AF$. Also, the rectangle $AFNL$ equals the square $FBKM$, that is, the rectangle on the whole and the segment AF equals the square on the complementary segment FB .

13.1.3. Books 3 and 4

The next two books take up topics familiar from high-school geometry: (a) circles, tangents, and secants and (b) inscribed and circumscribed polygons. Book 4 shows how to inscribe a regular pentagon in a circle (Proposition 11—this construction requires the construction given in Proposition 11 of Book 2 and shown in Fig. 13.2) and how to circumscribe a regular pentagon about a circle (Proposition 12), and then it reverses the figures and shows how to get the circles given the pentagon (Propositions 13 and 14). After the easy construction of a regular hexagon (Proposition 15), Euclid finishes off Book 4 with the construction of a regular pentakaidecagon (15-sided polygon, Proposition 16).

13.1.4. Books 5 and 6

The Eudoxan theory of geometric proportion is expounded in Book 5, which contains the construction of the mean proportional between two lines (Proposition 13 of Book 5). In Proposition 16 of Book 6, Euclid proves the all-important result that if four lines (line segments, as we would say) are in proportion, then the rectangle on the means equals the rectangle on the extremes. This theory is applied to solve the problems of application with defect and excess. A special case of the latter, in which it is required to construct a rectangle on a given line having area equal to the square on the line and with a square excess is the very famous Section (Proposition 30). Euclid phrases the problem as follows: *Divide a line into mean and extreme ratio*. This means to find a point on the line so that the whole line is to one part as that part is to the second part. The Pythagorean theorem is then generalized to cover not merely the squares on the sides of a right triangle, but any similar polygons on those sides (Proposition 31). The book finishes with the well-known statement that central and inscribed angles in a circle are proportional to the arcs they subtend.

13.1.5. Books 7–9

Euclid's exposition of Greek number theory in Books 7–9 was discussed in Chapter 9. Here, since irrationals cannot occur, the notion of proportion is redefined to eliminate the need for the Eudoxan technique.

13.1.6. Book 10

The tenth book is evidently placed so as to be an extension of the rational theory of proportion in the number-theory books to a more detailed study of the situation when incommensurable lines arise that can be handled with no techniques beyond square roots, what we would call quadratic irrationals. This book occupies fully one-fourth of the entire length of the *Elements*. For its sheer bulk, one would be inclined to consider it the most important of all the 13 books, yet its 115 propositions are among the least studied of all, principally because of their technical nature. The irrationals constructed in this book by taking square roots are needed in the theory developed in Book 13 for inscribing regular solids in a sphere (that is, finding the lengths of their sides knowing the radius of the sphere). The book begins with the operating principle of the method of exhaustion, also known as the principle of Archimedes. That is, if a quantity is continually cut in half, it will eventually become smaller than any preassigned quantity of the same type. The way to demonstrate incommensurability through the Euclidean algorithm then follows as Proposition 2: *If, when the smaller of two given quantities is continually subtracted from the larger, that which is left never divides evenly the one before it, the quantities are incommensurable.* We used this method of showing that the side and diagonal of a regular pentagon are incommensurable in Chapter 9.

13.1.7. Books 11–13

The basic concepts of solid geometry, namely planes, parallelepipeds, and pyramids, are introduced in Book 11. The theory of proportion for these solid figures and the regular solids (cube, tetrahedron, octahedron, dodecahedron, icosahedron) is developed in Book 12, which contains the theorem that circles are proportional to the squares on their diameters (Proposition 2). This result, as mentioned in Chapter 11, is needed to make Hippocrates' quadrature of a lune rigorous.

Book 12 continues the development of solid geometry by establishing the usual proportions and volume relations for solid figures; for example, a triangular prism can be divided by planes into three equal pyramids (Proposition 7), a cone is one-third of a cylinder on the same base, and similar cones and cylinders are proportional to the cubes of their linear dimensions, ending with the proof that spheres are proportional to the cubes on their diameters (Proposition 18). Archimedes (or someone who edited his works) credited these theorems to Eudoxus.

Book 13, the last book of the *Elements*, is devoted to the construction of the regular solids and the relation between their sides, diagonals, and apothems and the radius of the sphere in which they are inscribed. The last proposition (Proposition 18) sets out the sides of these regular solids and their ratios to one another. An informal discussion following this proposition concludes that there can be only five regular solids.

13.2. THE DATA

Euclid's *Elements* assume a certain familiarity with the principles of geometric reasoning, principles that are explained in more detail in the *Data*. The Greek name of this work (*Dedoména*) means [*Things That*] *Have Been Given*, just as *Data* does in Latin. The title is a good description of the contents, which discuss the conditions that determine various geometric objects uniquely. The propositions in this book can be interpreted in various

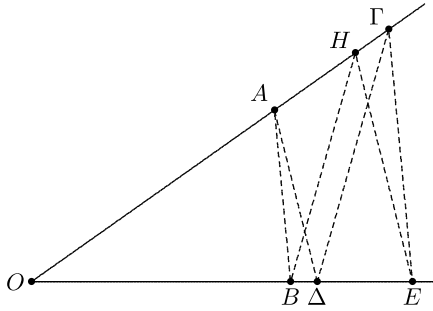


Figure 13.3. Proposition 81 of Euclid’s *Data*.

ways. Some can be looked at as *uniqueness* theorems. For example (Proposition 53), if the shapes—that is, the angles and ratios of the sides—are given for two polygons, and the ratio of the areas of the polygons is given, then the ratio of any side of one to any side of the other is given. Here, being given means being uniquely determined. Uniqueness is needed in proofs and constructions so that one can be sure that the result will be the same no matter what choices are made. It is an issue that arises frequently in modern mathematics, where operations on sets are defined by choosing representatives of the sets; when that is done, it is necessary to verify that the operation is *well-defined*, that is, independent of the choice made. An example of this need has already been cited in the previous chapter in connection with the definition of a composite ratio. In geometry we frequently say, “Let ABC be a triangle having the given properties *and having such-and-such a property*,” such as being located in a particular position. In such cases we need to be sure that the additional condition does not restrict the generality of the argument. In still another sense, this same proposition may guarantee that an explicit construction is *possible*, thereby removing the necessity of including it in the exposition of a theorem.

Other propositions assert that certain properties are *invariant*. For example (Proposition 81), when four lines $A, B, \Gamma,$ and Δ are given, and the line H is such that $\Delta : E = A : H$, where E is the fourth proportional to $A, B,$ and Γ , then $\Delta : \Gamma = B : H$. This proposition, illustrated in Fig. 13.3, is trivial to prove in modern notation. It also follows from Proposition 16 of Book 6 of the *Elements*, since $AE = B\Gamma = \Delta H$, that is, $A : B :: \Gamma : E$ and $\Delta : E :: AH$. This proposition is a lemma that can be useful in working out locus problems, which require finding a curve on which a point must lie if it satisfies certain prescribed conditions. Finally, a modern mathematician might interpret the assertion that an object is “given” as saying that the object “exists” and can be meaningfully talked about. To Euclid, that existence would mean that the object was explicitly constructible. (However, note that in two-dimensional figures, Euclid was inclined to gloss over these details, as in the proof of Proposition 2 of Book 12, exhibited in Chapter 12.)

PROBLEMS AND QUESTIONS

Mathematical Problems

- 13.1. Show how the construction of the Section (Fig. 13.3) can be interpreted as a problem in application with square excess. (See Fig. 13.4.)

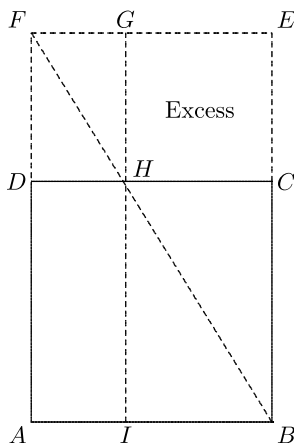


Figure 13.4. The Section as a problem in application with square excess. Rectangle $BEGI$, which is applied to the extension of BC , equals square $ABCD$ by Proposition 5 of Book 2. The “excess” $CEGH$ is a square by the construction of the Section.

- 13.2.** Use an argument similar to the argument in Chapter 9 showing that the side and diagonal of a pentagon are incommensurable to show that the side and diagonal of a square are incommensurable. That is, show that the Euclidean algorithm, when applied to the diagonal and side of a square, requires only two steps to produce the side and diagonal of a smaller square and hence can never produce an equal pair. [*Hint:* In Fig. 13.5, $AB = BC$, angle ABC is a right angle, AD is the bisector of angle CAB , and DE is drawn perpendicular to AC . Prove that $BD = DE$, $DE = EC$, and $AB = AE$. Then show that the Euclidean algorithm starting with the pair (AC, AB) leads first to the pair $(AB, EC) = (BC, BD)$ and then to the pair $(CD, BD) = (CD, DE)$, and these last two are the diagonal and side of a square.]
- 13.3.** The problem of constructing a rectangle of prescribed area on part of a given base a in such a way that the defect is a square is equivalent, when formulated as a problem about the lengths of the sides, to the problem of finding two numbers given their sum and product (the two numbers are the lengths of the sides of the rectangle). Stated in algebraic language, this problem asks for two numbers u and v (interpreted as lengths), such that $u + v = L$ and $uv = A$, where L is the given length and A the given area.

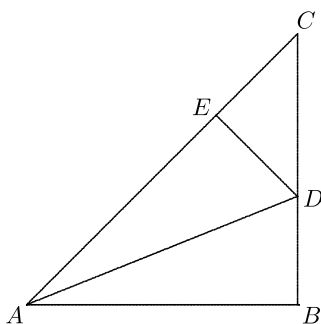


Figure 13.5. Diagonal and side of a square.

Show that this is equivalent to solving the quadratic equation $x^2 + A = Lx$. [Hint: If u and v are the roots of this equation, then $(x - u)(x - v) = 0$ when x is a root.] Since the largest rectangle that can be constructed on part of a line L so as to leave a square defect has area $L^2/4$ (the square on half of the line), the inequality $L^2 \geq 4A$ must hold. What does this condition say in terms of the algebra. [Hint: Look at the quadratic formula for solving the equation $x^2 - Lx + A = 0$.]

Historical Questions

- 13.4. Was the *Elements* an exposition of the most advanced mathematics of its time?
- 13.5. What does Euclid have to say in the *Elements* about the problems of squaring the circle, doubling the cube, and trisecting the angle?
- 13.6. Why does Euclid give two definitions of ratio and proportion, one in Book 6, the other in Book 7?

Questions for Reflection

- 13.7. If you were editing the *Elements*, how would you separate it into major parts. Which books belong in each?
- 13.8. Proposition 14 of Book 2 of Euclid shows how to construct a square equal in area to a rectangle. Since this construction is logically equivalent to constructing the mean proportional between two line segments, why does Euclid wait until Book 6, Proposition 13 to give the construction of the mean proportional?
- 13.9. Reflecting on the “geometric algebra” hypothesis, consider the plausibility of the following pseudo-syllogism: Mathematician A proved result B. Result B is logically equivalent to result C. Therefore mathematician A proved result C. Is this valid reasoning from the point of view of history? From the point of view of heritage? (Refer to Chapter 1 for an explanation of the difference between the two.)

Archimedes of Syracuse

Archimedes is one of a small number of mathematicians of antiquity of whose works we know more than a few fragments and of whose life we know more than the approximate time and place. The man indirectly responsible for his death, the Roman general Marcellus, is also indirectly responsible for the preservation of some of what we know about him. Archimedes lived in the Greek city of Syracuse, the name the Greeks seem to have used for the whole island of Sicily on which the city is located, during the third century BCE and is said by Plutarch to have been “a relative and a friend” of King Hieron II. Since Sicily lies nearly on a direct line between Carthage and Rome, it became embroiled in the Second Punic War (218–201 BCE). Marcellus took the city of Syracuse after a long siege, and Archimedes was killed by a Roman soldier in the chaos of the final fall of the city. In the course of writing a biography of Marcellus, Plutarch included some information on Archimedes.

According to Plutarch’s biography of Marcellus, the general was very upset that Archimedes had been killed and had his body buried in a suitably imposing tomb. When a nation is conquered, it often happens that the conquerors are insufficiently appreciative of its cultural achievements and the conquered nation is unable to preserve the relics of that culture. Such was the case with Archimedes. According to Eutocius, a biography of Archimedes was written by a certain Heracleides, who is mentioned in some of Archimedes’ letters. However, no copy of this biography is known to exist today. A century after Archimedes’ death, his tomb had fallen into neglect. In his *Tusculan Disputations*, written in 45 BCE, the famous Roman orator and statesman Cicero recalled his discovery of this tomb in 76 BCE.

When I was quaestor I tracked out [Archimedes’] grave, which was unknown to the Syracusans (as they totally denied its existence), and found it enclosed all round and covered with brambles and thickets; for I remembered certain doggerel lines inscribed, as I had heard, upon his tomb, which stated that a sphere along with a cylinder had been set up on the top of his grave. . . Slaves were sent in with sickles who cleared the ground of obstacles. . . So you see, one of the most famous cities of Greece. . . would have been ignorant of the tomb of its one most ingenious citizen, had not a man of Arpinum pointed it out.

During the Middle Ages, the tomb of Archimedes was lost again. In popular tradition, several tombs were erroneously believed to belong to Archimedes. However, the actual tomb

may have been rediscovered in 1957, during an excavation.¹ Since Syracuse was taken in 212 BCE and Archimedes was reported by the twelfth-century Byzantine writer Johannes Tzetzes to have been 75 years old at the time of his death, his dates are generally given as 287–212.

There are many legends connected with Archimedes, scattered among the various sources. Plutarch, for instance, says that Archimedes made many mechanical contrivances but generally despised such work in comparison with pure mathematical thought. Plutarch also reports three different stories of the death of Archimedes and tells us that Archimedes wished to have a sphere inscribed in a cylinder carved on his tombstone. The famous story that Archimedes ran naked through the streets shouting “Eureka!” (“I’ve got it!”) when he discovered the principle of specific gravity in the baths is reported by the Roman architect Vitruvius. Proclus gives another well-known anecdote: that Archimedes built a system of pulleys that enabled him (or King Hieron) single-handedly to pull a ship through the water. Finally, Plutarch and Pappus both quote Archimedes as saying in connection with his discovery of the principle of the lever that if there were another earth, he could move this one by standing on it (“δός μοι ποῦ στῶ καὶ κινῶ τῆν γῆν”).

14.1. THE WORKS OF ARCHIMEDES

With Archimedes we encounter the first author of a considerable body of original mathematical research that has been preserved to the present day. He was one of the most versatile, profound, creative, imaginative, rigorous, and influential mathematicians who ever lived. Ten of Archimedes’ treatises have come down to the present, along with a *Book of Lemmas* that seems to be Archimedean. Some of these works are prefaced by a “cover letter” intended to explain their contents to the person to whom Archimedes sent them. These correspondents of Archimedes were: Gelon, son of Hieron II and one of the kings of Syracuse during Archimedes’ life; Dositheus, a student of Archimedes’ student and close friend Conon; and Eratosthenes. Like the manuscripts of Euclid, all of the Archimedean manuscripts date from the ninth century or later. These manuscripts have been translated into English and published by various authors. A complete set of Medieval manuscripts of Archimedes’ work has been published by Marshall Clagett in the University of Wisconsin series on Medieval Science. In 1998, a palimpsest² of Archimedes’ work was sold at auction for \$2 million.

The 10 treatises referred to above are the following.

1. *On the Equilibrium of Planes*, Part I
2. *Quadrature of the Parabola*
3. *On the Equilibrium of Planes*, Part II

¹This claim was made by Professor Salvatore Ciancio (1965) on the basis of several criteria, including the location and date of the relics and a gold signet ring found in the crematory urn inside the tomb and bearing the ancient seal of the city of Alexandria. The sphere and cylinder mentioned by Cicero were not part of the find. The claim was contradicted at the time by the Curator of Antiquities in Syracuse Prof. Bernabò Brea. Another counterclaim is made by D. L. Simms in “The trail for Archimedes’ tomb,” *Journal of the Warburg and Courtauld Institute*, 53 (1990), pp. 281–286 (reference taken from the Worldwide Web). More information can be obtained at the address <http://www.mcs.drexel.edu/~crrres/Archimedes/contents.html>.

²That is, a book in which earlier work has been written and washed off so that new material could be entered in it.

4. *On the Sphere and the Cylinder*, Parts I and II
5. *On Spirals*
6. *On Conoids and Spheroids*
7. *On Floating Bodies*
8. *Measurement of a Circle*
9. *The Sand-reckoner*
10. *The Method*

References by Archimedes himself and other mathematicians tell of the existence of other works by Archimedes, of which no manuscripts are now known to exist. These include works on the theory of balances and levers, optics, the regular polyhedra, the calendar, and the construction of mechanical representations of the motion of heavenly bodies.

From this list we can see the versatility of Archimedes. His treatises on the equilibrium of planes and floating bodies contain principles that are now fundamental in mechanics and hydrostatics. The works on the quadrature of the parabola, conoids, and spheroids, the measurement of the circle, and the sphere and cylinder extend the theory of proportion, area, and volume found in Euclid for polyhedra and polygons to the more complicated figures bounded by curved lines and surfaces. The work on spirals introduces a new class of curves, and it develops the theory of length, area, and proportion for them.

Since we do not have space to discuss all of Archimedes' geometry, we shall confine the details of our discussion to what may be his greatest achievements: finding a planar region equal to the surface of a sphere and a polygonal region equal to a segment of a parabola. In addition, because of its impact on the issues involving proof that we have been discussing, we shall discuss his *Method* and show how he used it to discover certain results on quadrature.

14.2. THE SURFACE OF A SPHERE

Archimedes' two works on the sphere and cylinder were sent to Dositheus. In the letter accompanying the first of these, he gives some of the history of the problem. Archimedes considered his results on the sphere to be rigorously established, but he did have one regret. He wished he could have published them before Conon's death, "for he is the one we regard as most capable of understanding and rendering a proper judgment on them."

Archimedes sought a plane surface equal to the surface of a sphere by looking at a "hybrid figure," which curves in only one direction. (A sphere curves in every direction.) Specifically, he looked at the lateral surface of a frustum of a cone. In our terms, the area of a frustum of a cone with upper radius r , lower radius R , and side of slant height h is $\pi h(R + r)$. Archimedes, working in the Euclidean tradition, did not use formulas like this. Rather, he said that the frustum is equal to a disk whose radius is the mean proportional between the slant height and the sum of the two radii—that is, a disk whose radius is $\sqrt{h(R + r)}$. In our exposition, we shall use this fact in the following form: *A rectangle whose length is $R + r$ and whose width is h equals the square on the radius of a circle equal to the lateral surface of the frustum.* A conical frustum is shown in Fig. 14.1, and the simple way of getting its area is illustrated in Fig. 14.2. To save tedium, we will not repeat Archimedes' proof of this fact, but rather explain it in terms of modern geometry, using formulas.



Figure 14.1. Left: Frustum of a cone of slant height h and upper radius r . Right: Same frustum, turned upside down to exhibit the lower radius R .

When the frustum shown in Fig. 14.1 is cut open and laid flat, it occupies part of the annulus between two concentric circles of radii r' and R' , respectively. The portion α of this annulus occupied by the frustum is directly proportional to the portion of a complete circle the two boundaries of the frustum occupy. That portion is the same for both circles. The full circles would have circumference $2\pi r'$ and $2\pi R'$, respectively, whereas the portions of them corresponding to the boundaries of the frustum have length $2\pi r$ and $2\pi R$, respectively, where r and R are the radii of the boundary circles before the frustum was cut open. In other words, $\alpha = r/r' = R/R'$. Now the area of the full annulus, as we know, is $\pi(R'^2 - r'^2) = \pi(R' - r')(R' + r') = \pi h(R' + r')$. Hence the portion of it occupied by the cut-open frustum is $\pi h(\alpha R' + \alpha r') = \pi h(R + r)$. In other words, the radius of a circle whose area equals the lateral area of the frustum is $\sqrt{h(R + r)}$, as asserted. Notice that the formula works even when the “frustum” is a complete cone, that is, when $r = 0$. The area of a cone of base radius R and slant height h is πRh . (In the even more extreme case when the cone is flattened into a disk, we get $h = R$, and this same formula gives the area of the disk.)

This result can be applied to the figures generated by revolving a circle about a diameter with a regular $4n$ -sided polygon inscribed in it. Archimedes illustrated his argument with $n = 4$, but we shall illustrate it $n = 2$, that is, an inscribed octagon. The general idea is sufficiently clear from that case. Because all the right triangles in Fig. 14.3 are similar, we have $A'B : AB :: BE : AE :: B'E : EF :: CG : FG :: C'G : GH :: DI : HI :: D'I : IA'$.

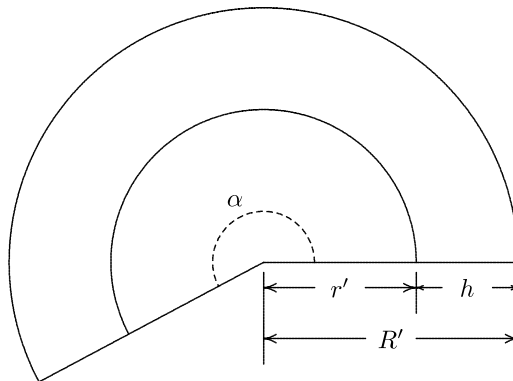


Figure 14.2. The frustum cut open. The angle α is to a complete rotation as the radius of the circular section at each point is to the slant distance to the apex of the cone. For the top and bottom circles of the frustum, those slant distances are shown here as r' and R' . Thus $R : R' :: r : r' :: \alpha : \text{complete rotation}$, where r and R are the radii of the upper and lower base of the frustum.

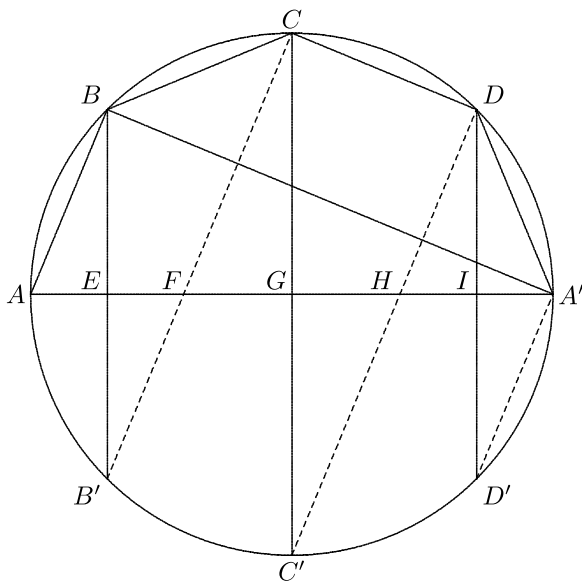


Figure 14.3. Finding the surface area of a sphere.

Since these ratios are all equal, we can add the first and second terms of any subset of them without changing the ratio. In particular, then

$$A'B : AB :: (BE + B'E + CG + C'G + DI + D'I) : (AE + EF + FG + GH + HI + IA').$$

Since this last term is just AA' (the diameter of the sphere), we get

$$A'B : AB :: (BE + (B'E + CG) + (C'G + DI) + D'I) : AA'.$$

Since all the chords in this figure are equal to AB , which is the slant height of the cones or frusta generated when the figure is revolved about the diameter AA' , each of the rectangles whose sides are AB and BE , AB and $B'E + CG$, AB and $C'G + DI$, and AB and $D'I$ is equal to the square whose side is the radius of a circle equal to corresponding cone or frustum. (The sums $BE + 0$, $B'E + CG$, $C'G + DI$, and $D'I + 0$ are the sums of the two radii of the frusta.)

At this point, we need a small fact that is easy to prove metrically, but more complicated to state in Euclid's language: *If S_1 and S_2 are squares whose sides are the radii of disks D_1 and D_2 respectively, and S_3 is a square such that $S_3 = S_1 + S_2$, then the side of square S_3 is the radius of a disk D_3 such that $D_3 = D_1 + D_2$.* This result is easily proved from the proportionality of disks and the squares on their radii (Proposition 2 of Book 12 of the *Elements*). When combined with the Pythagorean theorem, this result implies that the disk whose radius is the hypotenuse of a right triangle is the sum of the disks whose radii are the legs.

Thus, the fact (*Elements*, Book 6, Proposition 16) that when four lines are in proportion, the rectangle on the means equals the rectangle on the extremes means that the rectangle

on AB and the sum $BE + (B'E + CG) + (C'G + DI) + D'I$ equals the rectangle on $A'B$ and AA' . But the former, as just noted, is equal to the square on the radius of a disk equal to the sum of all these frusta. This is Proposition 22 of Archimedes' paper: *The rectangle on AA' and AB is the square on the radius of a circle equal to the sum of the (cones and) frusta generated by revolving the figure.*

The final step is to appeal to the method of exhaustion. If the side AB is chosen sufficiently small, $A'B$ can be made as close as we like to the diameter AA' , and the sum of the areas of the frusta can be made as close as we like to the surface of the sphere. Thus, as we would now put it, "in the limit" as the number of sides increases without bound, we find that the radius of a disk equal to the surface of the sphere is AA' , the diameter of the sphere. (Proposition 33): *The surface of any sphere is equal to four times the greatest circle in it.*

This achievement towers above anything found in any of Archimedes' predecessors. In order to get it, he had to make certain definitions about the area of a sphere, definitions that are in full agreement with intuition, but cannot be dispensed with even now. Once those definitions were made, he proved the result with full Euclidean rigor, leaving out no details. This argument is the only ancient expression for a plane region equal to the surface of a sphere that meets Euclidean standards of rigor.

Three remarks should be made on this proof. First, in view of the failure of efforts to square the circle, it seems that the later Greek mathematicians had two "standard" plane regions that could be used for comparing curved surfaces: the circle and the square. Archimedes expressed the surface of a sphere by finding a disk equal to it. Second, Archimedes could certainly have produced the "metric" version of this proof that is usually stated nowadays—namely that the area of a sphere of radius r is $4\pi r^2$ —since in his work on the measurement of a circle, he showed that a disk equals a right triangle whose legs are the radius and circumference of the disk, and also gave a numerical approximation to the ratio of the circumference and the diameter, which we express by the inequalities $3\frac{10}{71} < \pi < 3\frac{1}{7}$.³ Finally, Archimedes did not *discover* this theorem by Euclidean methods. He told how he came to discover it in his *Method*.

14.3. THE ARCHIMEDES PALIMPSEST

Early in the twentieth century the historian of mathematics J.L. Heiberg, reading in a bibliographical journal of 1899 the account of the discovery of a tenth-century manuscript with mathematical content, deduced from a few quotations that the manuscript was a copy of a work of Archimedes. In 1906 and 1908 he journeyed to Constantinople and established the text, as far as was possible. Attempts had been made to wash off the mathematical text during the Middle Ages so that the parchment could be used to write a book of prayers. The 177 pages of this manuscript contain parts of some of the works just discussed—it is the only source for the work on floating bodies—and a work called *Method*. The existence of such a work had been known because of the writings of commentators on Archimedes.

³These two results, interpreted in our language, imply that one-dimensional π —the ratio of circumference to diameter—and two-dimensional π —the ratio of a disk to the square on its radius—are the same number. But, being irrational, it was not a number to Archimedes.

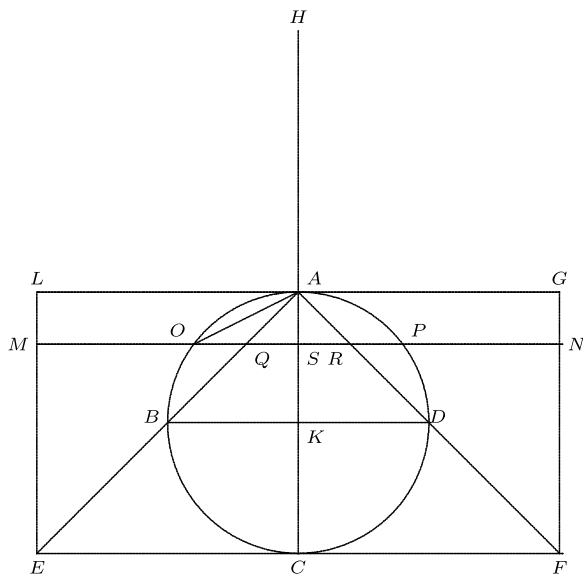


Figure 14.4. Volumes of sphere, cone, and cylinder.

14.3.1. The Method

There are quotations from the *Method* in the *Metrica*, of Heron of Alexandria, a work that was discovered in 1903. The *Method* had been sent to the astronomer Eratosthenes as a follow-up to a previous letter that had contained the statements of two theorems without proofs and a challenge to discover the proofs. Both of the theorems involve the volume and surface of solids of revolution. In contrast to his other work on this subject, however, Archimedes here makes free use of the principle now commonly known as *Cavalieri's principle*, which asserts that if the horizontal sections of two solids are in the same ratio at every elevation, then the volumes of those solids are in that same ratio. Archimedes' *Method* is a refinement of this principle, obtained by imagining the sections of a region balanced about a fulcrum. The reasoning is that if each pair of corresponding sections balance at distances a and b , then the bodies themselves will balance at these distances, and therefore, by Archimedes' principle of the lever, the area or volume of the two bodies must have the ratio $b : a$. Archimedes used this method to prove the following result:

A sphere is four times the cone with base equal to a great circle of the sphere and height equal to its radius. The cylinder with base equal to a great circle of the sphere and height equal to the diameter is half again as large as the sphere.

Archimedes' proof is based on Fig. 14.4. If this figure is revolved about the line CAH , the circle with center at K generates a sphere, the triangle AEF generates a cone, the rectangle $LGFE$ generates a cylinder, and each horizontal line such as MN generates a disk. The point A is the midpoint of CH . Archimedes shows that the area of the disk generated by revolving QR plus the area of the disk generated by revolving OP has the same ratio to the area of the disk generated by revolving MN that AS has to AH . It follows from his work on the equilibrium of planes that if the first two of these disks are hung at H ,⁴ they

⁴In looking at Fig. 14.4, you have to imagine that gravity is acting horizontally rather than vertically.

will balance the third disk about A as a fulcrum. Archimedes concluded that the sphere and cone together placed with their centers of gravity at H would balance (about the point A) the cylinder, whose center of gravity is at K .

Therefore,

$$HA : AK = (\text{cylinder}) : (\text{sphere} + \text{cone}).$$

But $HA = 2AK$. Therefore, the cylinder equals twice the sum of the sphere and the cone AEF . And since it is known that the cylinder is three times the cone AEF , it follows that the cone AEF is twice the sphere. But since $EF = 2BD$, cone AEF is eight times cone ABD , and the sphere is four times the cone ABD .

From this fact, Archimedes easily deduces the famous result allegedly depicted on his tombstone: *The cylinder circumscribed about a sphere equals the volume of the sphere plus the volume of a right circular cone inscribed in the cylinder.*

Having concluded the demonstration, Archimedes reveals that this method enabled him to find a planar region equal to the surface of a sphere. He writes

For I realized that just as every circle equals a triangle having as its base the circumference of the circle and altitude equal to the [distance] from the center to the circle [that is, the radius], in the same way every sphere is equal to a cone having as its base the surface of the sphere and altitude equal to the [distance] from the center to the sphere.

Thus, we now imagine two cones: One has a base circle equal to the surface of the sphere and height equal to the radius of the sphere, while the second is circumscribed about the sphere and hence has base circle equal to the equatorial circle of the sphere and height equal to twice its radius. Archimedes had established intuitively that the volume of the sphere was one-third of the former and two-thirds of the latter and, therefore, four-thirds of the cylinder obtained by taking the bottom half of the latter. But this last cylinder has the same height as the first, and therefore the volumes of the two are proportional to their bases. That means the base of the first cylinder (the area of the sphere) is four times the base of the last, which is the area of the equatorial circle. This is how Archimedes came to *discover* the result, which he then proved by the method of exhaustion (with a few reasonable assumptions about the approximation of areas). The method of exhaustion is very satisfying in settling an argument, but useless as a way of discovering the result. The *Method* shows us Archimedes' route to that discovery.

14.4. QUADRATURE OF THE PARABOLA

Archimedes used this method of imaginatively “balancing line segments” to show that the area of a parabolic segment is one-third larger than the largest triangle that can be inscribed in it. Having done that, he then proceeded to provide a rigorous proof of this fact using the method of exhaustion.

14.4.1. The Mechanical Quadrature

The mechanical proof is illustrated on the left side of Fig. 14.5, where K is the midpoint of the side BC of the triangle ABC and AC is the tangent to the parabola at A . Archimedes showed

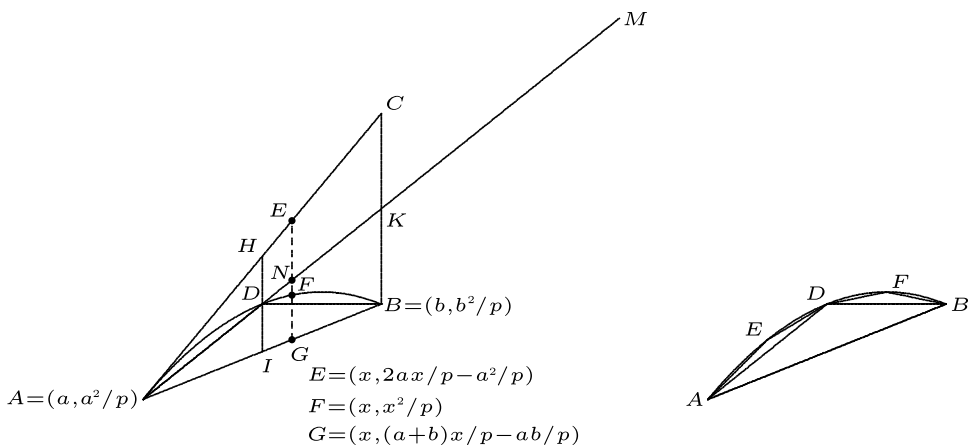


Figure 14.5. Left: Mechanical quadrature of a segment of the parabola $py = x^2$. Right: Rigorous quadrature based on the fact that $\triangle ABD$ is eight times as large as each of $\triangle AED$ and $\triangle BDF$.

that the line EG whose midpoint is at N will be exactly balanced about a fulcrum at K if the portion FG of the line is suspended at point M , where $AK = KM$. He then concluded that the whole triangle ABC in the position where it is would be exactly balanced by the parabolic segment hung at M . Since the center of gravity of the triangle lies one-third of the way from K to A , it follows that $\triangle ABC$ is three times the parabolic segment (which, for convenience, we shall refer to as S). But it is easily seen that $\triangle ABC$ is four times $\triangle ABD$, and hence S is one-third larger than $\triangle ABD$.

14.4.2. The Rigorous Quadrature

Although Archimedes willingly revealed the intuitive considerations that enabled him to solve the difficult problem of quadrature of a parabolic segment, he knew that this argument went beyond the pure methodology of the Euclidean tradition. He therefore followed this mechanical quadrature with a strictly rigorous proof, which we shall now describe.

Archimedes showed using properties of the parabola which we will not take the time to discuss, that if two more triangles AED and BFD are inscribed in the two sections cut off by sides AD and BD of the original triangle (see the right-hand side of Fig. 14.5, then these two triangles together equal exactly one-fourth of triangle ABD . Since adjoining the two new triangles removes more than half of the region between the segment and the triangle, it is clear that repeating this operation will eventually get a finite set of triangles which are together smaller than the parabolic segment, but differ from it by less than any specified magnitude.

Continuing in that way, doubling the number of new triangles at each step while reducing their total area by a factor of 4, he got what we would call an infinite (geometric) series for the magnitude of the parabolic segment:

$$\triangle ABD \times \left(1 + \frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \dots\right).$$

Since it would not have been acceptable in a proof to speak of the sum of infinitely many terms, Archimedes merely said that it was clear that the sum of the triangles at any stage of the operation was obviously not greater than S . That is, S could not be less than, for example,

$$\triangle ABD \times \left(1 + \frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \frac{1}{4^4} + \frac{1}{4^5}\right).$$

But, on the other hand, given any region U smaller than S , one could take enough triangles to get a sum of this form larger than U .

Archimedes observed that multiplying the last term by $4/3$ at any stage of the operation would cause the sum of the terms up to that stage to equal $4/3$. That is, this finite sum would collapse (“telescope,” we might say), if the last term were multiplied by $\frac{4}{3}$, since

$$\frac{1}{4^4} + \frac{1}{3 \times 4^4} = \frac{1}{3 \times 4^3}.$$

This relation causes the last two terms here to “merge” into $\frac{1}{3 \times 4^3}$. But then the last two terms of the altered sum merge similarly into $\frac{1}{3 \times 4^2}$, and this argument can be repeated until the whole sum reduces to $\frac{4}{3}$. Thus, he had shown that

$$\triangle ABD \times \left(1 + \frac{1}{4} + \cdots + \frac{1}{4^n}\right) = \frac{4}{3} \triangle ABD - \frac{1}{3 \cdot 4^n} \triangle ABD.$$

Hence the sum on the left is always less than $\frac{4}{3} \triangle ABD$, but given any region U smaller than $\frac{4}{3} \triangle ABD$, one could take enough triangles to get a sum of this form larger than U .

It is now clear that S must equal $\frac{4}{3} \triangle ABD$. If S were larger than $\frac{4}{3} \triangle ABD$, the first argument shows that we could take enough triangles to get a sum that is larger than $\frac{4}{3} \triangle ABD$, which contradicts the second argument. Exactly the same reasoning applies if we assume $\frac{4}{3} \triangle ABD$ is larger than S .

Observe that Archimedes has given a proof that is entirely finitistic and has not mentioned any infinite series here. He has proved two things: first, if $U > S$, then $U > \frac{4}{3} \triangle ABC$; second, if $U < S$, then $U < \frac{4}{3} \triangle ABC$. The assumption $\frac{4}{3} \triangle ABC < S$ would (by the second result) imply that $\frac{4}{3} \triangle ABC < \frac{4}{3} \triangle ABC$, which is absurd, and since the first result similarly rules out the possibility that $\frac{4}{3} \triangle ABC > S$, the only remaining possibility is that $\frac{4}{3} \triangle ABC = S$. In so doing, he has given exactly the kind of epsilon–delta proof that calculus students so often find mystifying. Newton once defended his formal rules in calculus (which we now establish rigorously by using the notion of a limit) by saying that it would be possible to justify them using the trichotomy of the ancient mathematicians but that there was no need to undergo such tedium. He must have had this method of exhaustion in mind.

PROBLEMS AND QUESTIONS

Mathematical Problems

- 14.1.** Use analytic geometry to justify the balancing in Archimedes' mechanical quadrature of the parabola. [Hint: Show that $EG : FG :: (b - a) : (b - x)$. Use the expressions shown for the points E , F , and G . Here the parabola is assumed to have the equation $py = x^2$, where p is negative, since the parabola opens downward, and we take the zero value of y to be the highest point on the parabola.]
- 14.2.** Prove that a disk D that equals the sum of disks D_1 of radius r_1 and D_2 of radius r_2 necessarily has radius r satisfying $r^2 = r_1^2 + r_2^2$. This is easy to do using the formula $A = \pi r^2$. Try to do it using only the fact that if P_1 and P_2 are similar polygons with a pair of corresponding sides equal to r_1 and r_2 , respectively, then $P_1 : P_2 :: D_1 : D_2$. (Use the result of Proposition 31 of Book 6 of the *Elements*, which asserts that the Pythagorean theorem holds for any similar polygons attached to the sides of a right triangle.)
- 14.3.** Give an alternative proof of Archimedes' result on the sphere and cylinder by showing (see Fig. 14.6) that $\overline{AB}^2 + \overline{AC}^2 = \overline{AB}^2 + \overline{OA}^2 = \overline{OB}^2 = \overline{AD}^2$. Hence the sum of the sections of the cone and sphere equals the section of the cylinder.

Historical Questions

- 14.4.** What advances in geometry, beyond the basic results found in the *Elements*, are due to Archimedes?
- 14.5.** What achievements of Archimedes show his versatility as a mathematician and scientist?

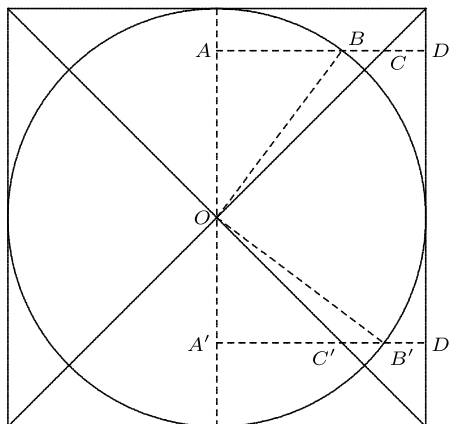


Figure 14.6. When the square circumscribed about the circle is revolved about its vertical midline (the line through A and A'), it generates the cylinder circumscribed about the sphere generated by the circle. The diagonals of the square generate a double-napped cone inscribed in the cylinder. The cylindrical section equals the sum of the sections of the cone and sphere.

- 14.6.** How did Archimedes discover that a sphere equals a cone whose base is the surface area of the sphere and whose height is the radius of the sphere?

Questions for Reflection

- 14.7.** Archimedes has a reputation as the greatest mathematician of antiquity and one of the three greatest of all time. What criteria can you imagine being applied to justify this judgment?
- 14.8.** Why, in your opinion, did Archimedes choose to give both a mechanical and a strict Euclidean proof of his quadrature results?
- 14.9.** It was mentioned above that some scholars now think Euclid was a contemporary of Archimedes rather than a predecessor. If that is so, how do you account for the fact that the *Elements* does not contain any material from Archimedes' treatises and that some of Archimedes works seem to refer to the *Elements*?

Apollonius of Perga

From what we have already seen of Greek geometry, we can understand how the study of the conic sections came to seem important. From commentators like Pappus, we know of treatises on the subject by Aristaeus, a contemporary of Euclid who is said to have written a book on *Solid Loci*, and by Euclid himself. We have also mentioned that Archimedes studied the conic sections. The only extensive treatise devoted just to conic sections that has survived, however, is that of Apollonius. Until recently, there was no adequate and accessible study of the whole treatise in English. The version most accessible was that of Heath, who said in his preface that writing his translation involved “the substitution of a new and uniform notation, the condensation of some propositions, the combination of two or more into one, some slight re-arrangements of order for the purpose of bringing together kindred propositions in cases where their separation was rather a matter of accident than indicative of design, and so on.” He also replaced Apollonius’ purely synthetic arguments with analytic arguments, based on the algebraic notation we are familiar with. All this labor has no doubt made Apollonius more readable. On the other hand, Apollonius’ work is no longer current research; and from the historian’s point of view, this kind of tinkering with the text only makes it harder to place the work in proper perspective. Nevertheless, one can fully understand the decision to use symbolic notation, since the mathematical language in which the original was couched was the cumbersome metric-free “synthetic” approach of Euclid in which the basic tools are lines and circles, and all relations must be reduced to proportions proved using something equivalent to the Eudoxan definition. A 1952 translation by R. Catesby Taliaferro of the first three books was included in the *Great Books of the Western World* series; it unfortunately went to the other extreme from the Heath translation and preserved the full obscurity of Apollonius’ original exposition. A translation of Books 5–7 (Toomer, 1990) at least made that portion of the work available to those like the present author, who could not read the Arabic in which the only extant manuscripts are written. Fortunately, all these gaps have now been filled in a thorough study of the entire work (Fried and Unguru, 2001).

In contrast to his older contemporary Archimedes, Apollonius remains a rather obscure figure. His dates can be determined from the commentary written on the *Conics* by Eutocius. Eutocius says that Apollonius lived in the time of the king Ptolemy Euergetes and defends him against a charge by Archimedes’ biographer Heraclides that Apollonius plagiarized results of Archimedes. Eutocius’ information places Apollonius in the second half of the third century BCE, perhaps a generation or so younger than Archimedes.

Pappus says that as a young man Apollonius studied at Alexandria, where he made the acquaintance of a certain Eudemus (probably not the student of Aristotle whose history of mathematics was used by Proclus). It is probably this Eudemus to whom Apollonius addresses himself in the preface to Book 1 of his treatise. From Apollonius' own words we know that he had been in Alexandria and in Perga, which had a library that rivaled the one in Alexandria. Eutocius reports an earlier writer, Geminus by name, as saying that Apollonius was called "the great geometer" by his contemporaries. He was highly esteemed as a mathematician by later mathematicians, as the quotations from his works by Ptolemy and Pappus attest. In Book 12 of the *Almagest*, Ptolemy attributes to Apollonius a geometric construction for locating the point at which a planet begins to undergo retrograde motion. From these later mathematicians we know the names of several works by Apollonius and have some idea of their contents. However, except for a few fragments that exist in Arabic translation, only two of his works survive to this day, and for them we are indebted to the Islamic mathematicians who continued to work on the problems that Apollonius considered important. Our present knowledge of Apollonius' *Cutting Off of a Ratio*, which contains geometric problems solvable by the methods of application with defect and excess, is based on an Arabic manuscript, no Greek manuscripts having survived. Of the eight books of Apollonius' *Conics*, only seven have survived in Arabic and only four in Greek. The astronomer Edmund Halley (1656–1743) published a Latin edition of all seven books in 1710. Halley also produced what Fried and Unguru (2001, p. 295) call "a reasonable and intelligent *partial* restoration" of Book 8, based on Apollonius' preface to Book 7, which he explains contains certain lemmas needed to prove what is in Book 8. As many people have pointed out, that statement does not necessarily cover the *entire* contents of Book 8; hence the use of the word *partial* by Fried and Unguru.

15.1. HISTORY OF THE CONICS

The evolution of the *Conics* was reported by Pappus five centuries after they were written in Book 7 of his *Collection*.

By supplementing Euclid's four books on the conics and adding four others Apollonius produced eight books on the conics. Aristaeus . . . and all those before Apollonius, called the three conic curves sections of acute-angled, right-angled, and obtuse-angled cones. But since all three curves can be produced by cutting any of these three cones, as Apollonius seems to have objected, [noting] that some others before him had discovered that what was called a section of an acute-angled cone could also be [a section of] a right- or obtuse-angled cone. . . changing the nomenclature, he named the so-called acute section an ellipse, the right section a parabola, and the obtuse section a hyperbola.

In a preface addressed to the aforementioned Eudemus, Apollonius lists the important results of his work: the description of the sections, the properties of the figures relating to their diameters, axes, and asymptotes, things necessary for analyzing problems to see what data permit a solution, and the three- and four-line locus. He continues:

The third book contains many remarkable theorems of use for the construction of solid loci and for distinguishing when problems have a solution, of which the greatest part and the most beautiful are new. And when we had grasped these, we knew that the three-line and four-line

locus had not been constructed by Euclid, but only a chance part of it and that not very happily. For it was not possible for this construction to be completed without the additional things found by us.

We have space to discuss only the definition and construction of the conic sections and the four-line locus problem, which Apollonius mentions in the passage just quoted.

15.2. CONTENTS OF THE *CONICS*

The earlier use of conic sections had been restricted to cutting cones with a plane perpendicular to a generator. As we saw in our earlier discussion, this kind of section is easy to analyze and convenient in the applications for which it was intended. In fact, only one kind of hyperbola, the rectangular, is needed for duplicating the cube and trisecting the angle. The properties of a general section of a general cone were not discussed. Also, it was considered a demerit that the properties of these plane curves had to be derived from three-dimensional figures. Apollonius set out to remove these gaps in the theory.

First it was necessary to define a cone as the figure generated by moving a line around a circle while one of its points, called the *apex* and lying outside the plane of the circle, remains fixed. Next, it was necessary to classify all the sections of a cone that happen to be circles. Obviously, those sections include all sections by planes parallel to the plane of the generating circle (Book 1, Proposition 4). Surprisingly, there is another class of sections that are also circles, called *subcontrary* sections. Once the circles are excluded, the remaining sections must be parabolas, hyperbolas, and ellipses.

We shall give some details of Apollonius' construction of the ellipse and then briefly indicate how the same procedure applies to the other conic sections. Consider the section of a cone shown in Fig. 15.1, made by a plane cutting all the generators of the cone on the same side of its apex. This condition is equivalent to saying that the cutting intersects both sides of the axial triangle (see Fig. 6 of Chapter 11). Apollonius proved that there is a certain line (EH in the figure), which he called *the* [up]right side (or *perpendicular side* or *vertical side*), now known by its Latin name *latus rectum*, such that the square on the ordinate from any point of the section to its axis equals the rectangle applied to the *latus rectum* with width equal to the abscissa and whose defect on the *latus rectum* is similar to the rectangle formed by the axis and the *latus rectum*. He gave a rule, too complicated to go into here, for constructing the *latus rectum*. This line characterized the shape of the curve. Because of its connection with the problem of application with defect, he called the resulting conic section an *ellipse*. Similar connections with the problems of application and application with excess, respectively, arise in Apollonius' construction of the parabola and hyperbola. These connections motivated the names he gave to these curves.

In Fig. 15.1, where the *latus rectum* is the line EH , the locus condition¹ is that the square on the ordinate LM equals the rectangle on EO and EM , that is, $\overline{LM}^2 = \overline{EO} \cdot \overline{EM}$. The

¹The Latin word *locus* is the equivalent of the Greek word *tópos*, from which our word *topology* comes. Both mean *place*. The Greek mathematicians had to imagine a cone generated by a line with one of its points fixed moving around a circle. A locus was thought of as the path followed by a moving point. Modern mathematics has replaced the concept of a locus by the concept of a set, meaning the points satisfying a certain condition. This concept is a static one, not the kinematic picture imagined by the Greeks. But it is more realistic, since a set may be disconnected and hence difficult to picture as the path of a moving point.

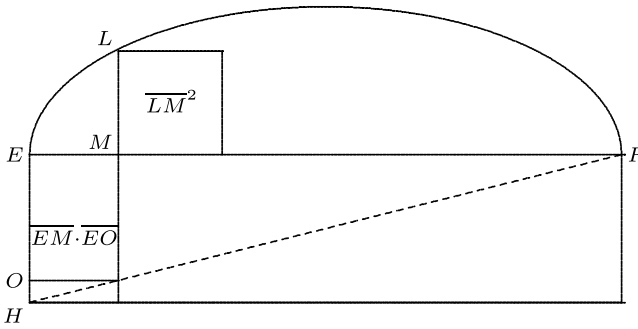


Figure 15.1. Apollonius’ construction of the ellipse with latus rectum $p = EH$. Given $x = EM$ and $y = LM$, these two lines are connected by the relation $\overline{LM}^2 = \overline{EM} \cdot \overline{EO} = \overline{EH} \cdot \overline{EM} - \overline{OH} \cdot \overline{EM}$. Since OH is proportional to EM , this says $y^2 = px - kx^2$, where k is the slope of the diagonal of the rectangle whose sides are the axis EF and the latus rectum EH .

reason for the term *ellipse* is that the rectangle applied to the latus rectum with area equal to the square on the ordinate and width equal to the abscissa leaves a defect of prescribed shape (the shape of the rectangle whose sides are the axis and the latus rectum) on the remainder of the latus rectum.

In one sense, this locus definition for an ellipse is not far removed from what we now think of as the equation of the ellipse, but that small gap was unbridgeable in Apollonius’ time. We shall digress briefly to “translate” this language to its modern algebraic equivalent, again warning the reader that Apollonius was certainly *not* thinking of the figure this way. If we write $LM = y$ and $EM = x$ in Fig. 15.1 (so that we are essentially taking rectangular coordinates with origin at E), we see that Apollonius is claiming that $y^2 = x \cdot EO$. Now, however, $EO = EH - OH$, and EH is constant, while OH is directly proportional to EM , that is, to x . Specifically, the ratio of OH to EM is the same as the ratio of the latus rectum EH to the axis EF . It follows that an ellipse is uniquely determined by the latus rectum and its major axis. Thus, if we write $OH = kx$ —a crucial step that Apollonius could not take, since he did not have the concept of a dimensionless constant of proportionality—and denote the latus rectum EH by p , we find that Apollonius’ locus condition can be stated as the equation $y^2 = px - kx^2$. Here k is the slope of the dashed line HF . By completing the square on x , transposing terms, and dividing by the constant term, we can bring this equation into what we now call the standard form for an ellipse with center at $(a, 0)$:

$$\frac{(x - a)^2}{a^2} + \frac{y^2}{b^2} = 1,$$

where $a = p/(2k)$ and $b = p/(2\sqrt{k})$. In this notation, the latus rectum p is $2b^2/a$. Apollonius, however, did *not* have the concept of an equation nor the symbolic algebraic notation we now use, and if he had known about these things, he would still have lacked the letter k used above as a constant of proportionality. These “missing” pieces gave his work on conics a ponderous character with which most mathematicians today have little patience. That is why Heath’s translation of the *Conics* looks more like a textbook of analytic geometry than an ancient Greek treatise translated into English.

Apollonius’ constructions of the parabola and hyperbola also depend on the latus rectum. A parabola is completely determined by its latus rectum and the locus condition that the

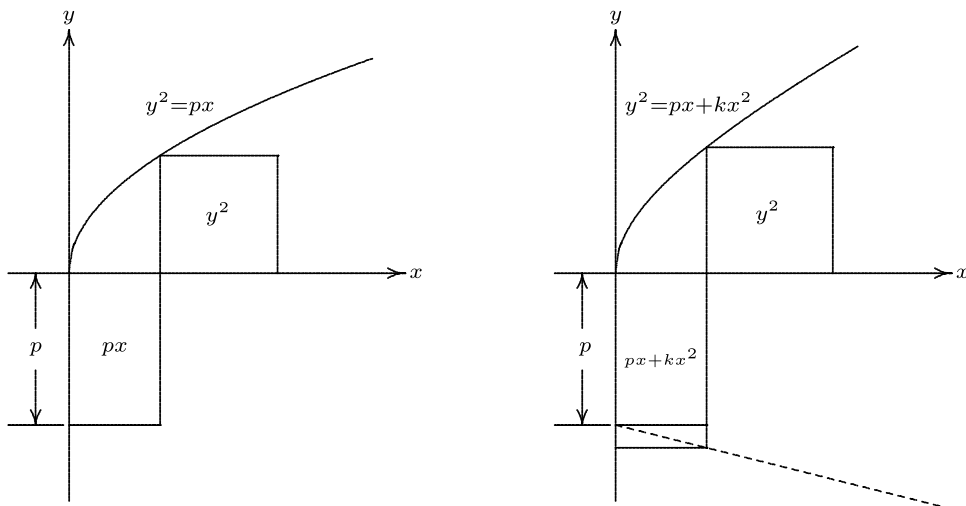


Figure 15.2. Left: The parabola with latus rectum p . The condition for a point to be on the locus is that the square on its ordinate (y) equal the rectangle on its abscissa (x) and the latus rectum (p), that is $y^2 = px$. Right: A hyperbola with latus rectum p . Here $y^2 = px + kx^2$, where the oblique dashed line has slope $-k$.

square on the ordinate equals the rectangle on the abscissa and the latus rectum ($y^2 = px$). In other words, the rectangle on the latus rectum whose width is the abscissa is exactly equal to the square on the ordinate, with no excess or defect. For a hyperbola, $y^2 = px + kx^2$, so that the hyperbola is determined by the latus rectum p and the constant k , which is the negative of the slope of the dashed line in the right-hand drawing in Fig. 15.2. In this case, the rectangle having a side along the latus rectum, width equal to the abscissa, and area equal to the square on the ordinate has length that exceeds the latus rectum, creating an “excess” rectangle whose shape is the same for all points on the hyperbola. The now-standard form for this equation is

$$\frac{(x - a)^2}{a^2} - \frac{y^2}{b^2} = 1,$$

and the latus rectum p is once again $2b^2/a$.

Apollonius was the first to take account of the fact that a plane whose intersection with a cone is a hyperbola must cut both nappes of the cone. He regarded the two branches as two hyperbolas, referring to them as “opposites” and using the term *hyperbola* for either branch. For the hyperbola, Apollonius proved the existence of *asymptotes*—that is, a pair of lines through the center that never meet the hyperbola but such that any line through the center passing into the region containing the hyperbola does meet the hyperbola. The word *asymptote* means literally *not falling together*—that is, not intersecting. For the hyperbola shown on the right-hand side of Fig. 15.2, the asymptotes are the two lines $y = p/(2\sqrt{k}) \pm \sqrt{k}x$.

With these new characterizations of the three conic sections, it becomes possible to discard the cone itself. Once the latus rectum and the shape of the excess or defect (measured in our terms by the constant of proportionality that we denoted by k) are given, the locus

condition defining the curve is determined. It makes no reference to anything outside the plane of the curve itself. The original cone is like the scaffolding around a building, which is removed after the construction is complete. With these curves now defined as plane loci, their properties can then be developed using Euclid's plane geometry. Apollonius proceeds to do so.

15.2.1. Properties of the Conic Sections

Books 1 and 2 of the *Conics* are occupied with finding the proportions among line segments cut off by chords and tangents in conic sections, the analogs of results on circles in Books 3 and 4 of the *Elements*. These constructions involve finding the tangents to the curves satisfying various supplementary conditions such as being parallel to a given line. Fried and Unguru (2001, Chapter 7) argue that these analogies probably guided Apollonius in his choice of material.

15.3. FOCI AND THE THREE- AND FOUR-LINE LOCUS

We are nowadays accustomed to constructing the conic sections using the focus–directrix property, so that it comes as a surprise that the original expert on the subject does not seem to recognize the importance of the foci. He never mentions the focus of a parabola, and for the ellipse and hyperbola he refers to these points only as “the points arising out of the application.” The “application” he has in mind is explained in Book 3. Propositions 48 and 52 of Book 3 read as follows:

(Proposition 48) *If in an ellipse a rectangle equal to the fourth part of the figure is applied from both sides to the major axis and deficient by a square figure, and from the points resulting from the application straight lines are drawn to the ellipse, the lines will make equal angles with the tangent at that point.*

(Proposition 52) *If in an ellipse a rectangle equal to the fourth part of the figure is applied from both sides to the major axis and deficient by a square figure, and from the points resulting from the application straight lines are drawn to the ellipse, the two lines will be equal to the axis.*

The “figure” referred to is the rectangle whose sides are the major axis of the ellipse and the latus rectum. In Fig. 15.3 the points F_1 and F_2 must be chosen on the major axis AB so

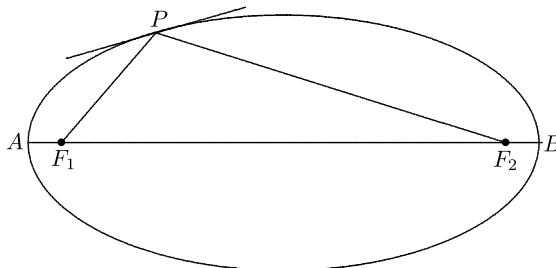


Figure 15.3. Focal properties of an ellipse.

that the rectangle on AF_1 and F_1B and the rectangle on AF_2 and BF_2 both equal one-fourth of the area of the rectangle formed by the whole axis AB and the latus rectum p .

Proposition 48 expresses the focal property of these two points: A light ray emanating from one will be reflected to the other. Proposition 52 is the *string property* that characterizes the ellipse as the locus of points such that the sum of the distances to the foci is constant. These are just two of the theorems Apollonius called “strange and beautiful.” Apollonius makes little use of these properties, however, and does not discuss the use of the string property to draw an ellipse.

A very influential part of the *Conics* consists of Propositions 54–56 of Book 3, which contain the theorems that Apollonius claimed (in his cover letter) would provide a solution to the three- and four-line locus problems. Both in their own time and because of their subsequent influence during the seventeenth century (when analytic geometry was being created), the three- and four-line locus problems have been of great importance for the development of mathematics. These propositions involve the proportions among pieces of chords inscribed in a conic section. Three propositions are needed because the hyperbola requires two separate statements according as the points involved lie on the same or opposite branches of the hyperbola.

We limit ourselves to stating the four-line locus problem and illustrating it. The data for the problem are four lines, which for definiteness we suppose to intersect two at a time, and four given angles, one corresponding to each line. The problem requires the locus of points P such that if lines are drawn from P to the four lines, each making the corresponding angle with the given line (for simplicity all shown as right angles in Fig. 15.4), the rectangle on two of the lines will have a constant ratio to the rectangle on the other two. The solution is in general a pair of conics.

The origin of this kind of problem may lie in the problem of two mean proportionals, which was solved by drawing fixed reference lines and finding the loci of points satisfying a condition resembling the condition here. In that problem, the square on the line drawn perpendicular to one reference line equals the rectangle on a fixed line and the line drawn to the other reference line. The commentary on this problem by Pappus, who mentioned that Apollonius had left a great deal unfinished in this area, inspired Fermat and Descartes to take up the implied challenge and solve the problem completely. Descartes offered his success in solving the locus problem to any number of lines as proof of the value of his analytic method in geometry.

PROBLEMS AND QUESTIONS

Mathematical Problems

- 15.1.** The string property of an ellipse is illustrated in Fig. 15.3. It implies that all broken lines starting at one focus, going to any point on the ellipse, and then going to the other focus have the same total length. Taking for granted that an ellipse is a convex figure, you may assume that the tangent to an ellipse at any point lies entirely outside the ellipse, except for the point of tangency itself. Use this fact to prove the reflection property of the ellipse. You will need to establish that if two points D and E are on the same side of a given line MN , then the shortest path from D to a point Q on the line and thence to E is the one for which the lines DQ and QE make equal angles with the line MN . This theorem is illustrated in Fig. 15.5.

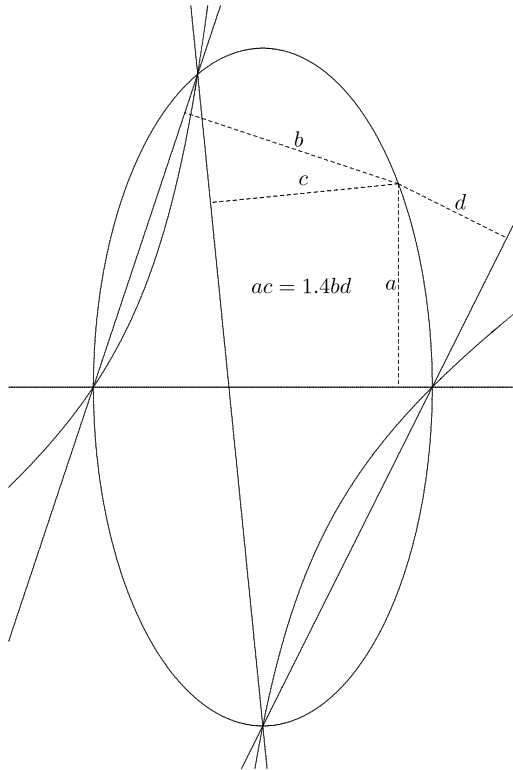


Figure 15.4. The four-line locus. If a point moves so that the product of its distances to two lines bears a constant ratio to the product of its distances to two other lines, it must move in a conic. In this illustration, two conics satisfy the condition: one an ellipse, the other a hyperbola.

- 15.2. Show from Apollonius' definition of the foci that the product of the distances from each focus to the ends of the major axis of an ellipse equals the square on half of the minor axis.
- 15.3. We have seen that the three- and four-line locus problems have conic sections as their solutions. State and solve the two-line locus problem. You may use modern analytic geometry and assume that the two lines are the x axis and the line $y = ax$.

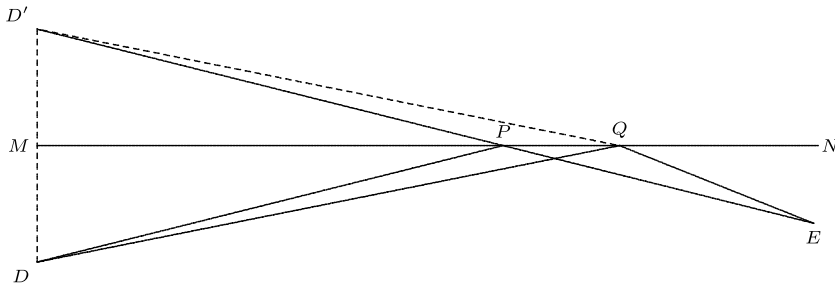


Figure 15.5. Shortest path meeting a line.

The locus is the set of points whose distances to these two lines have a given ratio. What curve is this? (The distance from a point (u, v) to the line whose equation is $ax + by = c$ is $\frac{|au + bv - c|}{\sqrt{a^2 + b^2}}$.)

Historical Questions

- 15.4. How much of the treatise on conics by Apollonius has been preserved, and in what form?
- 15.5. On what basis can we conjecture what was in the missing Book 8 of the *Conics*?
- 15.6. Why did Apollonius rename the conic sections?

Questions for Reflection

- 15.7. As we have seen, Apollonius was aware of the string property of ellipses, yet he did not mention that this property could be used to draw an ellipse. Do you think that he did not *notice* this fact, or did he omit to mention it because he considered it unimportant, or for some other reason?
- 15.8. Is the apparent generality of Apollonius' statement of the three-line locus problem, in which arbitrary angles can be prescribed at which lines are drawn from the locus to the fixed lines, really more general than the particular case in which all the angles are right angles? Observe that the ratio of a line from a point P to line l making a fixed angle θ with the line l bears a constant ratio to the line segment from P perpendicular to l . How would a particular locus problem be altered if recast in terms of the perpendicular distances to the same lines?
- 15.9. A circle can be regarded as a special case of an ellipse. What is the *latus rectum* of a circle? (Consider the expression given for the latus rectum in terms of the semi-axes of the ellipse.)

Hellenistic and Roman Geometry

No sensible person would attempt to study modern geometry using Euclidean methods. The difficulty with such an approach is not only that the kind of proof required is a very long and tedious way of proving even the simplest results. The basic objects underlying almost all of the geometry of Euclid, Archimedes, and Apollonius were few and simple, being formed from straight lines, planes, circles, and combinations of them. The conic sections are already near the limit of tolerable complexity that can be generated from these tools. The small number of more complicated curves considered in ancient times, such as the spiral of Archimedes, the quadratrix of Hippias, and the conchoid of Nicomedes were created by introducing the concept of motion into the basically static geometry of Euclid's *Elements*. They were pictured as the path of a point moving under simple conditions that combined straight-line and circular motion. Curved three-dimensional figures such as spheres and cones were likewise pictured as the result of translating or revolving lines or circles. In contrast, the kinds of curves studied in modern mathematics, such as the graphs of polynomials of degree three and higher or transcendental functions such as the logarithm or the sine, are impossible to analyze using these tools. The Euclidean methodology set limits to the growth of geometry, and those limits were nearly reached by the end of the third century BCE. Still, a few later mathematicians attempted to go beyond beyond the achievements of Archimedes and Apollonius, and they produced some good work over the next few centuries.

16.1. ZENODORUS

The astronomer Zenodorus lived in Athens in the century following Apollonius. Although his exact dates are not known, he is mentioned by the late third-century mathematician Diocles in his book *On Burning Mirrors*¹ and by Theon of Alexandria in his commentary on Ptolemy's *Almagest*. According to Theon, Zenodorus wrote *On Isoperimetric Figures*, in which he proved four theorems: (1) If two regular polygons have the same perimeter, the one with the larger number of sides encloses the larger area; (2) a circle encloses a larger area than any regular polygon whose perimeter equals its circumference; (3) of all polygons with a given number of sides and perimeter, the regular polygon is the largest; (4) of all closed surfaces with a given area, the sphere encloses the largest volume. With

¹Arabic manuscripts of this work have revealed that Diocles and Zenodorus actually collaborated (Toomer, 1976).

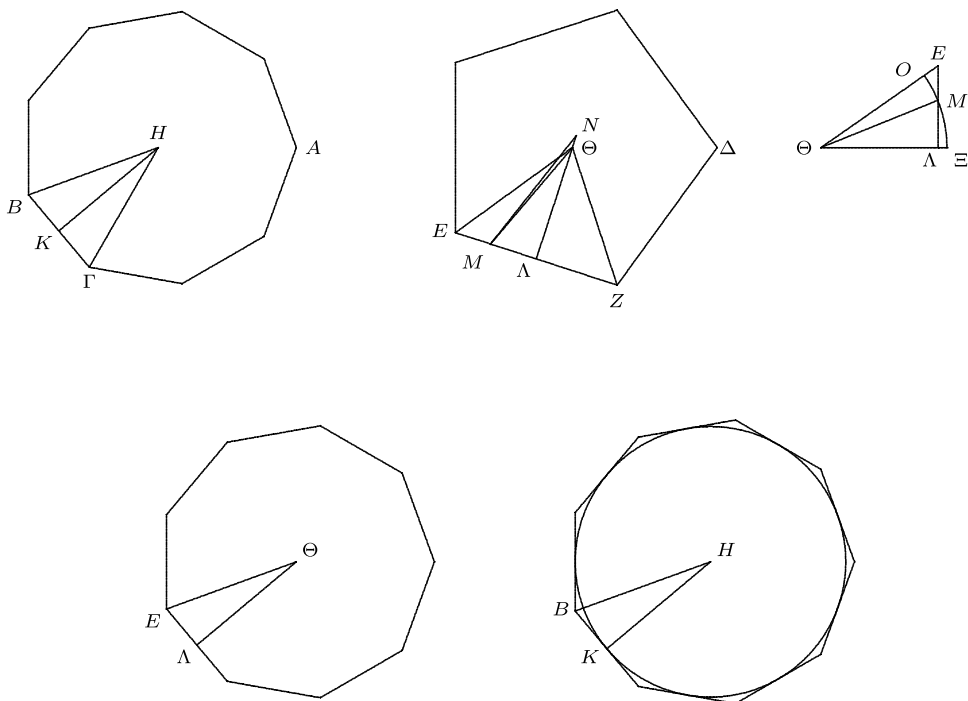


Figure 16.1. Two theorems of Zenodorus. Top: When two regular polygons have the same perimeter, the one with the larger number of sides is larger. Bottom: A circle is larger than a regular polygon whose perimeter equals the circumference of the circle.

the machinery inherited from Euclidean geometry, Zenodorus could not have hoped for any result more general than these. Let us examine his proof of the first two, as reported by Theon.

In Fig. 16.1, let $AB\Gamma$ and ΔEZ be two regular polygons having the same perimeter, with $AB\Gamma$ having more sides than ΔEZ . Let H and Θ be the centers of these polygons, and draw the lines from the centers to two adjacent vertices and their midpoints, getting triangles $B\Gamma H$ and $EZ\Theta$ and the perpendicular bisectors of their bases HK and $\Theta\Lambda$. Then, since the two polygons have the same perimeter but $AB\Gamma$ has more sides, BK is shorter than $E\Lambda$. Mark off M on $E\Lambda$ so that $M\Lambda = BK$. Then if P is the common perimeter, we have $E\Lambda : P :: \angle E\Theta\Lambda : 4 \text{ right angles}$ and $P : BK :: 4 \text{ right angles} : \angle BHK$. By composition then (see Subsection 12.2.3 of Chapter 12), $E\Lambda : BK :: \angle E\Theta\Lambda : \angle BHK$, and therefore $E\Lambda : M\Lambda :: \angle E\Theta\Lambda : \angle BHK$. But, Zenodorus claimed, the ratio $E\Lambda : M\Lambda$ is larger than the ratio $\angle E\Theta\Lambda : \angle M\Theta\Lambda$, asking to postpone the proof until later. Granting that lemma, he said, the ratio $\angle E\Theta\Lambda : \angle BHK$ is larger than the ratio $\angle E\Theta\Lambda : \angle M\Theta\Lambda$, and therefore $\angle BHK$ is smaller than $\angle M\Theta\Lambda$. It then follows that the complementary angles $\angle HBK$ and $\angle \Theta M\Lambda$ satisfy the reverse inequality. Hence, copying $\angle HBK$ at M so that one side is along $M\Lambda$, we find that the other side intersects the extension of $\Lambda\Theta$ at a point N beyond Θ . Then, since triangles BHK and MNA are congruent by angle-side-angle, it follows that $HK = NA > \Theta\Lambda$. But the areas of the two polygons are $\frac{1}{2}HK \cdot P$ and $\frac{1}{2}\Theta\Lambda \cdot P$, and therefore $AB\Gamma$ is the larger of the two.

The proof that the ratio $EA : MA$ is larger than the ratio $\angle E\Theta\Lambda : \angle M\Theta\Lambda$ was given by Euclid in his *Optics*, Proposition 8. But Theon does not cite Euclid in his quotation of Zenodorus. He gives the proof himself, implying that Zenodorus did likewise. The proof is shown on the top right in Fig. 16.1, where the circular arc ΞMO has been drawn through M with Θ as center. Since the ratio $\triangle E\Theta M : \text{sector } O\Theta M$ is larger than the ratio $\triangle M\Theta\Lambda : \text{sector } M\Theta\Xi$ (the first triangle is larger than its sector, the second is smaller), it follows, interchanging means, that $\triangle E\Theta M : \triangle M\Theta\Lambda > \text{sector } O\Theta M : \text{sector } M\Theta\Xi$. But $\triangle E\Theta M : \triangle M\Theta\Lambda :: EM : M\Lambda$, since the two triangles have the same altitude $\Theta\Lambda$ measured from the base line $EM\Lambda$. And $\text{sector } O\Theta M : \text{sector } M\Theta\Xi :: \angle E\Theta M : \angle M\Theta\Lambda$. Therefore, $EM : M\Lambda$ is larger than the ratio $\angle E\Theta M : \angle M\Theta\Lambda$, and it then follows that $EA : MA$ is larger than $\angle E\Theta\Lambda : \angle M\Theta\Lambda$. (See the explanation of the addition of ratios in Subsection 12.2.3 of Chapter 12.)

Zenodorus' proof that a circle is larger than a regular polygon whose perimeter equals the circumference of the circle is shown at the bottom of Fig. 16.1. Given such a polygon and circle, circumscribe a similar polygon around the circle. Since this polygon is "convex on the outside," as Archimedes said in his treatise on the sphere and cylinder, it can be assumed longer than the circumference. (Both Archimedes and Zenodorus recognized that this was an assumption that they could not prove; Zenodorus cited Archimedes as having assumed this result.) That means the circumscribed polygon is larger than the original polygon since it has a larger perimeter. But then by similarity, HK is larger than $\Theta\Lambda$. Since a circle equals half of the rectangle whose sides are its circumference and radius (proved by Archimedes), while a regular polygon is half of the rectangle whose sides are its perimeter and its apothem,² it follows that the circle is larger.

16.2. THE PARALLEL POSTULATE

We saw in Chapter 12 that there was a debate about the theory of parallel lines in Plato's Academy, as we infer from the writing of Aristotle. This debate was not ended by Euclid's decision to include a parallel postulate explicitly in the *Elements*. This foundational issue was discussed at length by the Stoic philosopher Geminus, whose dates are a subject of disagreement among experts, but who probably lived sometime between 50 BCE and 50 CE. Geminus wrote an encyclopedic work on mathematics, which has been entirely lost, except for certain passages quoted by Proclus, Eutocius, and others. Proclus said that the parallel postulate should be completely written out of the list of postulates, since it is really a theorem. The asymptotes of hyperbolas provided the model on which he reasoned that converging is not the same thing as intersecting. But still he thought that such behavior was impossible for straight lines. He claimed that a line that intersected one of two parallel lines must intersect the other,³ and he reports a proof of Geminus that assumes in many places that certain lines drawn will intersect, not realizing that by doing so he was already assuming the parallel postulate.

²An apothem—not to be confused with an apothegm—is the line from the center of a polygon perpendicular to a side. In this case, the apothem is $\Theta\Lambda$.

³This assertion is an *assumption* equivalent to the parallel postulate and obviously equivalent to the form of the postulate commonly used nowadays, known as Playfair's axiom, after the Scottish geometer John Playfair (1748–1819): *Through a given point not on a line, only one parallel can be drawn to the line.*

Proclus also reports an attempt by Ptolemy to prove the postulate by arguing that a pair of lines could not be parallel on one side of a transversal “rather than” on the other side. (Proclus did not approve of this argument.) But the assumption that parallel lines are symmetric under reflection through a common transversal that is perpendicular to one of them is a Euclidean *theorem* that does not extend to non-Euclidean geometry. These early attempts to prove the parallel postulate began the process of unearthing more and more plausible alternatives to the postulate, but of course did not lead to a proof of it.

16.3. HERON

We have already noted some of the restrictions that Euclid imposed on plane geometry, one of them being that lines are not associated with numbers. After Apollonius, however, the metric aspects of geometry began to resurface in the work of later writers. One of these writers was Heron (ca. 10–ca. 75), who wrote on mechanics. He probably lived in Alexandria. Pappus discusses his work in Book 8 of his *Collection*. Heron’s geometry is much more concerned with measurement than was the geometry of Euclid. The change of interest in the direction of measurement and numerical procedures signaled by his *Metrica* is shown vividly by his repeated use (130 times, to be exact) of the word *area* (*embadón*), a word never once used by Euclid, Archimedes, or Apollonius.⁴ There is a difference in point of view between saying that two plane figures are equal and saying that they have the same area. The first statement is geometrical and is the stronger of the two. The second is purely numerical and does not necessarily imply the first. Heron discusses ways of finding the areas of triangles from their sides. After giving several examples of triangles that are either integer-sided right triangles or can be decomposed into such triangles by an altitude, such as the triangle with sides of length 13, 14, 15, which is divided into a 5–12–13 triangle and a 9–12–15 triangle by the altitude to the side of length 14, he gives “a direct method by which the area of a triangle can be found without first finding its altitude.” He gave as an example a triangle whose sides were 7, 8, and 9 units. His prescription was: Add 9 and 8 and 7, getting 24. Take half of this, getting 12. Subtract 7 units from this, leaving 5. Then subtract 8 from 12, leaving 4. Finally, subtract 9, leaving 3. Multiply 12 by 5, getting 60. Multiply this by 4, getting 240. Multiply this by 3, getting 720. Take the square root of this, and that will be the area of the triangle. He went on to explain that since 720 is not a square, it will be necessary to approximate, starting from the nearest square number, 729.

This result seems anomalous in Greek geometry, since Heron is talking about multiplying an area by an area. That is probably why he emphasizes that his results are numerical rather than geometric. His proof is based on Fig. 16.2, in which one superfluous line has been omitted to streamline it. In the following proof, some rewording has been done to accommodate this minor modification of the figure. This figure shows an arbitrary triangle $AB\Gamma$ with its inscribed circle, and the radii of that circle to the points of contact $H\Delta$,

⁴Reporting (in his commentary on Ptolemy’s *Almagest*) on Archimedes’ *Measurement of a Circle*, however, Theon of Alexandria did use this word to describe what Archimedes did; but that usage was anachronistic. In his work on the sphere, for example, Archimedes referred to its *surface* (*epipháneia*), not its *area*. On the other hand, Dijksterhuis (1956, pp. 412–413) reports the Arabic mathematician al-Biruni as having said that “Heron’s formula” is really due to Archimedes. Considering the contrast in style between the proof and the applications, it does appear plausible that Heron learned the proof from Archimedes. Heath (1921, Vol. 2, p. 322) endorses this assertion unequivocally.

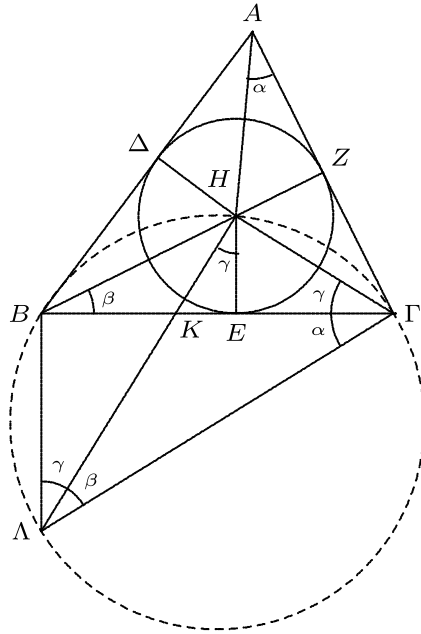


Figure 16.2. Proof of Heron’s method of computing the area of a triangle from the lengths of its sides.

HE , and HZ drawn. These three radii are all congruent, and each is perpendicular to the corresponding side. Denote their common length by r . The lines HA , HB , and $H\Gamma$ from the center of the inscribed circle to the vertex are the bisectors of the angles. We have marked one of each congruent pair at each vertex as α , β , and γ . It is then clear that $\alpha + \beta + \gamma$ is half the sum of the angles of the triangle, that is, it is equal to a right angle.

Because of the way the triangle is partitioned into three triangles $A\Gamma H$, ABH , and $B\Gamma H$, we can see that its area is $\frac{1}{2}HZ \cdot A\Gamma + \frac{1}{2}HE \cdot B\Gamma + \frac{1}{2}H\Delta \cdot AB$, which is easily seen to be $r \frac{AB+B\Gamma+\Gamma A}{2} = r\Sigma$, where Σ is half the perimeter of the triangle. There are many ways different ways of expressing Σ as a sum of lines, among them $\Sigma = \overline{B\Gamma} + \overline{AZ} = \overline{AB} + \overline{E\Gamma} = \overline{A\Gamma} + \overline{BE}$. We shall use all three of these expressions below.

Heron claims that the area is numerically $\sqrt{\Sigma(\Sigma - AB)(\Sigma - A\Gamma)(\Sigma - B\Gamma)}$, which, using these three expressions for Σ , we can write as $\sqrt{\Sigma \cdot \overline{E\Gamma} \cdot \overline{BE} \cdot \overline{AZ}}$.

To see why this expression represents the area of the triangle, draw lines $B\Lambda$ and $H\Lambda$ from B and H perpendicular, respectively, to $B\Gamma$ and $H\Gamma$ and intersecting at the point Λ . The quadrilateral $\Lambda B\Gamma H$ is cyclic—that is, can be inscribed in a circle. This fact holds because the two right triangles $\Lambda B\Gamma$ and $\Lambda H\Gamma$ have $\Lambda\Gamma$ as a common hypotenuse. Since the hypotenuse of a right triangle is a diameter of the circumscribed circle, that circumscribed circle is the same for both. Then, since $\angle H\Gamma B$ and $\angle B\Lambda H$ are both inscribed in the same arc \widehat{BH} , it follows that angle $B\Lambda H = \gamma$. Similarly, since $\angle H\Lambda\Gamma$ and $\angle H\Gamma B$ are both inscribed in the same arc $\widehat{H\Gamma}$, we have $\angle H\Lambda\Gamma = \beta$. Adding, we find that $\angle B\Lambda\Gamma = \beta + \gamma$, and therefore the complementary angle $B\Gamma\Lambda$ equals α . Finally, since both HE and $B\Lambda$ are perpendicular to $B\Gamma$, we have $\angle AHE = \angle B\Lambda H = \gamma$. We therefore have two pairs of similar right triangles: $\Delta B\Lambda$ and ΔAZH , and $\Delta B\Lambda K$ and ΔEHK .

From these two similar triangles, we obtain the fundamental results that $\frac{\overline{B\Gamma}}{\overline{B\Lambda}} = \frac{\overline{AZ}}{\overline{HZ}}$ and $\frac{\overline{B\Lambda}}{\overline{BK}} = \frac{\overline{HE}}{\overline{KE}} = \frac{\overline{HZ}}{\overline{KE}}$. By composition, $\frac{\overline{B\Gamma}}{\overline{BK}} = \frac{\overline{AZ}}{\overline{KE}}$, that is, $\overline{B\Gamma} \cdot \overline{KE} = \overline{BK} \cdot \overline{AZ}$. Also, since HE is the altitude to the hypotenuse of the right triangle $KH\Gamma$, we have $\overline{HE}^2 = \overline{KE} \cdot \overline{E\Gamma}$. These last two relations are the key to proving the result.

It is now possible to obtain the required expression for the area by plodding through the following sequence of equalities, each of which follows trivially from the one above.

$$\begin{aligned}\overline{BE} \cdot \overline{AZ} &= \overline{BK} \cdot \overline{AZ} + \overline{KE} \cdot \overline{AZ}, \\ \overline{BE} \cdot \overline{AZ} &= \overline{KE} \cdot \overline{B\Gamma} + \overline{KE} \cdot \overline{AZ}, \\ \overline{BE} \cdot \overline{AZ} &= \overline{KE} \cdot \Sigma, \\ \overline{E\Gamma} \cdot \overline{BE} \cdot \overline{AZ} &= \overline{E\Gamma} \cdot \overline{KE} \cdot \Sigma, \\ &= \overline{HE}^2 \cdot \Sigma, \\ \Sigma \cdot \overline{E\Gamma} \cdot \overline{BE} \cdot \overline{AZ} &= \overline{HE}^2 \cdot \Sigma^2. \\ \sqrt{\Sigma(\Sigma - \overline{AB})(\Sigma - \overline{A\Gamma})(\Sigma - \overline{B\Gamma})} &= \overline{HE} \cdot \Sigma.\end{aligned}$$

The result is therefore proved.

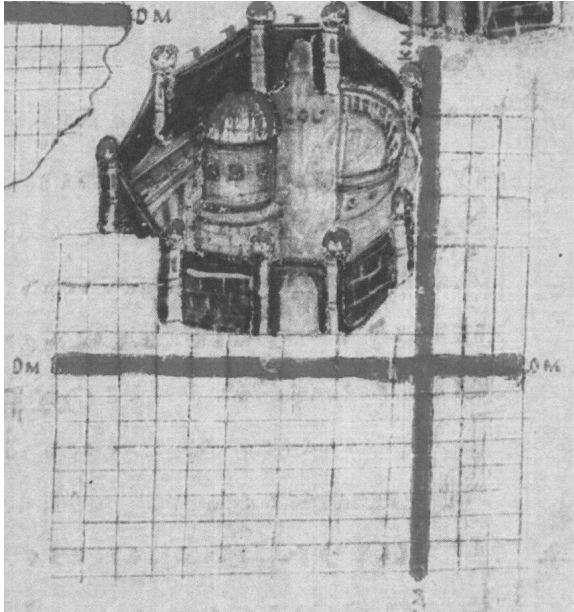
16.4. ROMAN CIVIL ENGINEERING

Dilke (1985, pp. 88–90) describes the use of geometry in Roman civil engineering as follows. The center of a Roman village would be at the intersection of two perpendicular roads, a (usually) north–south road called the *kardo maximus* (literally, the *main hinge*) and an east–west road called *decumanus maximus*, the *main tenth*. Lots were laid out in blocks (*insulae*) called *hundredths* (*centuriae*), each block being assigned a pair of numbers, telling how many units it was *dextra decumani* (right of the *decumanus*, that is, north⁵) or *sinistra decumani* (left of the *decumanus*, that is, south) and how many units it was *ultra kardinem* (beyond the *kardo*, that is, west) or *citra kardinem* (within the *kardo*, that is, east).

A collection of Roman writings on surveying was collected, translated into German, and published in Berlin in the middle of the nineteenth century. This two-volume work bears the title *Corpus Agrimensorum Romanorum*, the word *agrimensor* (field measurer) being the Latin name for a surveyor, as already noted in Chapter 7. A medieval town laid out in accordance with the scheme just described is shown in that work. Looking at it, one cannot help thinking of a rectangular coordinate system, regarding the *kardo* and *decumanus*, with the coordinatization of the *centuriae*, as prefigurations of our concept of coordinate axes. The spherical equivalent was used by Ptolemy, as we shall see in the next chapter, and his latitude and longitude did influence the development of analytic geometry.

Among the agrimensores was one named M. Iunius Nipsus, a second-century surveyor, who, according to Dilke (1985, p. 99), gives the following directions for measuring the width of a river (Fig. 16.3).

⁵This orientation presumes the map user is looking west along the *decumanus maximus*. Often, the town forum would be located at the intersection of the two main roads.



Portion of a Roman town, from the *Corpus Agrimensorum Romanorum*, Hans Butzmann, ed., W. Sijthoff, Leyden, 1970 (Cod. Guelf. 36.23 Aug 2^o. fol. 63v). Copyright © Herzog August Bibliothek Wolfenbüttel.

You mark the point C on the opposite bank from B (a part of the procedure Nipsus neglects to mention until later), continue the straight line CB to some convenient point A , lay down the crossroads sign at A , and then move along the direction perpendicular to AC until you reach a point G , where you erect a pole, then continue on to D so that $GD = AG$. You then move away from D along the direction perpendicular to AD until you see G and C in a straight line from the point H . Since the triangles AGC and DGH are congruent (by angle–side–angle), it follows that $CB = CA - AB = HD - AB$.

For this procedure to work in practice, it is necessary to have an accessible and level piece of land covering the lines shown as AD and DH . If the river is large, such a stretch of land may not exist, since the river banks are likely to be hilly. In its neglect of similar triangles, this method seems a large step backward in applied geometry.

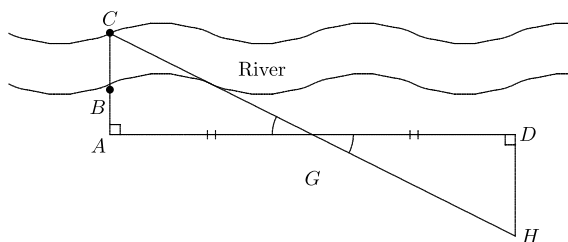


Figure 16.3. Nipsus' method of computing the width of a river.

PROBLEMS AND QUESTIONS

Mathematical Problems

- 16.1.** Find the area of a triangle whose sides are 24 cm, 37 cm, and 43 cm.
- 16.2.** Express the area of a regular n -gon in terms of its perimeter P , using trigonometry. What happens to this area as n tends to infinity?
- 16.3.** Suppose that four squares A , B , C , and D are in proportion, that is, $A : B :: C : D$. Let their sides be respectively a , b , c , and d . Prove that the sides are also in proportion, that is, $a : b :: c : d$, using the Eudoxan definition of proportion. [*Hint:* Let m and n be any two integers. You need to show that if $ma > nb$, then $mc > nd$. Assuming $ma > nb$, prove that $m^2A > n^2B$. Why does $m^2A > n^2B$ imply $m^2C > n^2D$, and why does this last inequality imply $mc > nd$?]

Historical Questions

- 16.4.** What new directions did Greek geometry explore after the treatises of Archimedes and Apollonius?
- 16.5.** What isoperimetric inequalities were stated and proved by Zenodorus?
- 16.6.** What relation did Heron establish between the lengths of the three sides of a triangle and its area?

Questions for Reflection

- 16.7.** Which of the two writers discussed above, Heron and Zenodorus, worked more in the spirit of Euclid's *Elements*?
- 16.8.** How did Heron regard the ratio of two lines? Did he handle it as Eudoxus recommended?
- 16.9.** Why did the elaborate system of Euclidean geometry apparently play no role in the magnificent engineering feats (roads, aqueducts, and the like) of the Romans?

Ptolemy's Geography and Astronomy

The path away from the metric-free geometry of Euclid, Archimedes, and Apollonius was opened by Heron, Ptolemy, and other geometers who lived during the early centuries of the Roman Empire. Ptolemy's *Almagest* is an elegant arithmetization of some basic Euclidean geometry applied to astronomy. As in the work of Heron, proportions provided the key to arithmetizing triangles and circles in a way that made a computational geometric model of the motions of the heavenly bodies feasible. In the *Almagest*, computations using a table of chords are combined with rigorous geometric demonstration of the relations involved. Ptolemy (ca. 85–ca. 165), whose very name shows his Alexandrian origins even though he lived in Rome, shows that he is well acquainted with the geometry and astronomy of his day. But he studied the earth as well as the sky, and his contribution to geography is also a large one, and also geometric, although less computational than his astronomy. We shall begin with the geography, for which an annotated translation exists (Berggren and Jones, 2000).

17.1. GEOGRAPHY

Ptolemy was one of the first scholars to look at the problem of representing large portions of the earth's surface on a flat map. His data, understandably very inaccurate from the modern point of view, came from his predecessors, including the astronomers Eratosthenes (276–194) and Hipparchus (190–120) and the geographers Strabo (ca. 64 BCE–24 CE) and Marinus of Tyre (70–130), whom he followed in using the now-familiar lines of latitude and longitude. These lines have the advantage of being perpendicular to one another, but the disadvantage that the parallels of latitude are of different sizes. Hence a degree of longitude stands for different east–west distances at different latitudes.

As an empirical science, geography depends intimately on these two coordinates, whose empirical foundations are very different. Latitude is relatively easy to determine. If you look at any known star when it is directly above your local north–south line, you can measure its altitude above the southernmost point on the horizon. Then if you add 90° and subtract the known declination of that star on the celestial sphere, the difference will be your geographical latitude.¹ Longitude is much more difficult to determine, since the same stars will pass overhead at the same local time on a given day at all points having the same latitude. To determine longitude, it is necessary to single out one meridian, called the *prime*

¹For the definitions of declination and local altitude, see Section 17.2 below.

meridian, for use as the origin. By universal convention, that meridian is the one through Greenwich, England. You can work out your longitude east or west of Greenwich provided you know what time it is in Greenwich at any moment of your choosing. The difference between your local solar time and the local solar time at Greenwich is your longitude. (Each hour of time difference represents 15 degrees of longitude.) In these days of global positioning systems, terrestrial longitudes are known precisely, and there is no reason to doubt the data. But in the days before the instant communication of radio, and in the absence of a clock that would keep perfect time while being transported over a considerable portion of the earth's surface, longitude was very difficult to determine. What was needed was an event that could be observed from all over the earth simultaneously, to be followed by a comparison of the local times at which they occur. One such event occurred at the time of the Battle of Arbela, which was mentioned in Chapter 11. According to Pliny the Elder (23–79) in his *Natural History*, Book 2, Chapter 72)²:

We are told that at the time of the famous victory of Alexander the Great, at Arbela, the moon was eclipsed at the second hour of the night, while, in Sicily, the moon was rising at the same hour.

This would mean that Sicily is about 30° west of Arbela, and in fact Sicily straddles the meridian at 14° E while Arbela is at 44° E.

Such events are rare, and comparisons of them are hard to coordinate. The search continued for a large celestial clock. The phases of the moon would seem to be an obvious one, which everyone on the same side of the earth can observe at once. However, they change too slowly to allow precise measurement. When he discovered four moons of Jupiter, Galileo realized that their configuration, which changed fairly rapidly, could be used as the clock he wanted. Once again, however, measurements could not be made with sufficient precision to be of any use. Not until a durable and accurate spring-wound clock was created could this problem be effectively solved. Ptolemy was forced to rely on travel times over east–west routes to determine relative longitudes.

Ptolemy assigned latitudes to the inhabited spots that he knew about by computing the length of daylight on the longest day of the year. This computational procedure is described in Book 2, Chapter 6 of the *Almagest*, where Ptolemy describes the latitudes at which the longest day lasts $12\frac{1}{4}$ hours, $12\frac{1}{2}$ hours, and so on up to 18 hours, then at half-hour intervals up to 20 hours, and finally at 1-hour intervals up to 24. Although he knew theoretically what the Arctic Circle is, he didn't know of anyone living north of it, and took the northernmost location on the maps in his *Geography* to be Thoulē, described by the historian Polybius around 150 BCE as an island six days sail north of Britain that had been discovered by the merchant–explorer Pytheas (380–310) of Masillia (Marseille) some two centuries earlier.³ It has been suggested that Thoulē is the Shetland Islands (part of Scotland since 1471), located between 60° and 61° north; that is just a few degrees south of the Arctic Circle, which is at 66° 30'. It is also sometimes said to be Iceland, which is on the Arctic Circle, but west of Britain as well as north. Whatever it was, Ptolemy assigned it a latitude of 63°, although he said in the *Almagest* that some “Scythians” (Scandinavians and Slavs) lived still farther north at $64\frac{1}{2}$ °. Ptolemy did know of people living south of the equator and

²Arbela is modern Erbil, Iraq, but the battle took place some 80 km distant from it. At that battle in 331 BCE, Alexander defeated the Persian King Darius and effectively put an end to the Persian Empire.

³The Latin idiom *ultima Thule* means roughly *the last extremity*.

took account of places as far south as Agisymba (Ethiopia) and the promontory of Prasmus (perhaps Cabo Delgado in Mozambique, which is 14° south). Ptolemy placed it $12^\circ 30'$ south of the Equator. The extreme southern limit of his map was the circle $16^\circ 25'$ south of the equator, which he called “anti-Meróē,” since Meróē (a city on the Nile River in southern Egypt) was $16^\circ 25'$ north.

Since he knew only the geography of what is now Europe, Africa, and Asia, he did not need 360° of longitude. He took his westernmost point to be the Blessed Islands (possibly the Canary Islands, at 17° west). That was his prime meridian, and he measured longitude out to 180° eastward from there, to the Sēres,⁴ the Chinese (Sínai), and “Kattígara.” According to Dilke (1985, p. 81), “Kattígara” may refer to Hanoi. Actually, the east–west span from the Canary Islands to Shanghai (about 123° east) is only 140° of longitude. Ptolemy’s inaccuracy is due partly to unreliable reports of distances over trade routes and partly to his decision to accept 500 stades, about 92.5 km—a stade is generally taken to have been 185 m—as the length of a degree of latitude. The true distance is about 600 stades, or 111 km.⁵ We are not concerned with the units in Ptolemy’s geography, however, only with its mathematical aspects.

The problem Ptolemy faced was to draw a flat map of the earth’s surface spanning 180° of longitude and about 80° degrees of latitude, from $16^\circ 25'$ south to 63° north. Ptolemy described three methods of doing this, the first of which we shall now discuss. The latitude and longitude coordinates of the inhabited world (*oikuménē*) known to Ptolemy represent a rectangle whose width is $\frac{5}{9}$ of its length. Ptolemy did not represent parallels of latitude as straight lines; he drew them as arcs of concentric circles while keeping the meridians of longitude as straight lines emanating from the common center, representing the North Pole. Thus, he mapped this portion of the earth into the portion of a sector of a disk bounded by two radii and the arcs they cut off on two circles concentric with the disk. As shown in Fig. 17.1, the first problem was to decide which radii and which circles are to form these boundaries. Ptolemy recognized that it would be impossible in such a map to place all the parallels of latitude at the correct distances from one another and still get their lengths in proportion. He decided to keep his northernmost parallel, through Thoúlē, in proportion to the parallel through the equator. That meant these two arcs should have a ratio of about 9:20—more precisely, $\cos(63^\circ)$ in our terms, which is 0.45399. Since there would be 63 equal divisions between that parallel and the equator, he needed the upper radius x to satisfy $x : (x + 63) :: 9 : 20$. Solving this proportion is not hard, and one finds that $x = 52$, to the nearest integer. The next task was to decide on the angular opening. For this principle he decided, like Marinus of Tyre, to get the parallel of latitude through Rhodes in the correct proportion. Since Rhodes is at 36° latitude, the length of half of the parallel of latitude through it amounts to about $\frac{4}{5}$ of the 180° arc of a great circle, which is about 145° . Since the radius of Rhodes must be 79 (27 great-circle degrees more than the radius of Thoúlē), he needed the opening angle of the sectors θ to satisfy $\theta : 180^\circ :: 146 : 79\pi$, so that $\theta \approx 106^\circ$. After that, he inserted meridians of longitude every one-third of an hour of longitude (5°) fanning out from the North Pole to the Equator.

⁴The Sēres were a Hindu people known to the Greeks from the silk trade.

⁵One measurement of the length of a degree available to Ptolemy was that of Eratosthenes, who used shadow lengths at the summer solstice in Syene and Alexandria, Egypt to conclude (correctly) that Alexandria is about 7° north of Syene. (It is actually about 3° west as well.) Eratosthenes gave the distance between the two cities as 5000 stades. The actual distance is about 730 km. These figures are inconsistent with the length of a stade given above and the number of stades in a degree, but Ptolemy had other sources to reconcile with Eratosthenes.

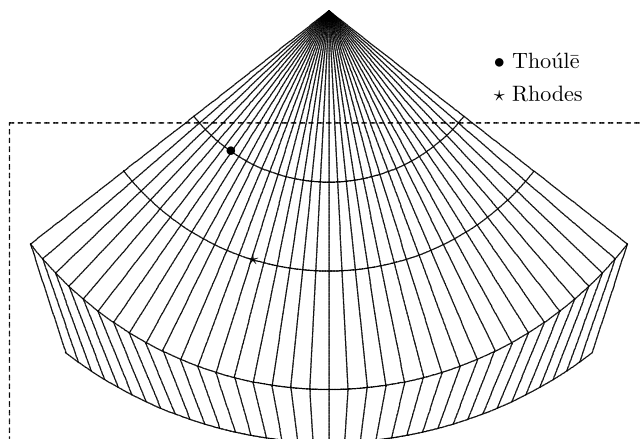


Figure 17.1. Ptolemy's first method of mapping.

Continuing to draw the parallels of latitude in the same way for points south of the Equator would lead to serious distortion, since the circles in the sector continue to increase as the distance south of the north pole increases, while the actual parallels on the earth begin to decrease at that point. The simplest solution to that problem was to let the southernmost parallel at $16^{\circ} 25'$ south have its actual length, then join the meridians through that parallel by straight lines to the points where they intersect the equator. Once that decision was made, he was ready to draw the map on a rectangular sheet of paper. He gave instructions for how to do that: Begin with a rectangle that is approximately twice as long as it is wide, draw the perpendicular bisector of the horizontal (long) sides, and extend it above the upper edge so that the portion above that edge and the whole bisector are in the ratio $34^{\circ} : 131^{\circ}, 25'$. In that way, the 106° arc through Thoúlē will begin and end just slightly above the upper edge of the rectangle, while the lowest point of the map will be at the foot of the bisector, being about 80 units below the lowest point on the parallel of Thoúlē, as indicated by the dashed line in Fig. 17.1.

Although at first sight, this way of mapping seems to resemble a conical projection, it is *not* that, since it preserves north–south distances. It does a tolerably good job of mapping the parts of the world for which Ptolemy had reliable data.

17.2. ASTRONOMY

Ptolemy's astronomy, for which there are good sources in English (Toomer 1984a,b, Jones 1990) was much more geometrical than his geography. It established metrical relations among chords and arcs of circles by means of which it was possible to give latitude and longitude coordinates (what are now called *declination* and *right ascension*, respectively) for all the stars and planets (including the sun and moon among the latter). These coordinates were imposed on what is called the *celestial sphere*, which is a representation of all the stars as if they were stuck on a sphere whose center was at the center of the earth. It is shown in Fig. 17.2. The circle labeled *ecliptic* in that figure is the path that the sun follows as it moves through the fixed stars, making one circuit per year. It crosses the celestial equator moving from south to north on March 20 (rarely, on March 21). That intersection of the ecliptic and equator is called the *vernal equinox*. It provides a natural prime meridian

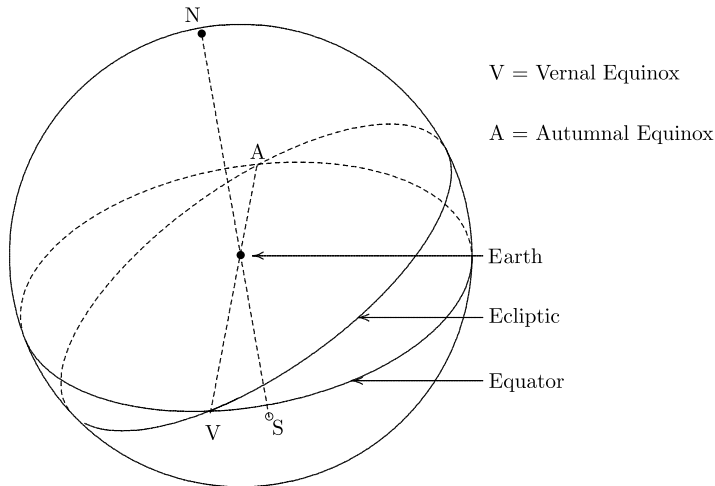


Figure 17.2. The celestial sphere.

on the celestial sphere, from which right ascension can be measured. (There is no such natural prime meridian on the earth, and the one actually used—through Greenwich—is a human convention.) Since it was impossible in Ptolemy’s day to tell the distances to the fixed stars, what we call the radial coordinate in three-dimensional spherical coordinates was irrelevant. Using this celestial sphere, with the vernal equinox serving as origin, one could assign permanent locations to all the fixed stars, leaving only seven celestial bodies known to Ptolemy (sun, moon, Mercury, Venus, Mars, Jupiter, and Saturn) as “wandering” stars whose coordinates were constantly changing among the fixed stars. The problem for geocentric astronomy was to express that wandering as a combination of simple circular motions.

As with his geography, Ptolemy benefited from data assembled over a long period of time, namely observations of the positions of various planets, times of eclipses, and other celestial phenomena, made at various places in the Mediterranean world, including Mesopotamia, over the preceding 800 years. To fit all these data to observation, he used a system known as epicycles and/or eccentrics. An epicycle is a uniform motion about the center of a circle that is itself moving uniformly around a second circle, called the *deferent*. An eccentric is a uniform motion in a circle, but observed from a point not at the center of the circle. Either of these devices can be used to account for the observed variable speeds of a heavenly body around the celestial sphere.

17.2.1. Epicycles and Eccentrics

The fact that a uniform motion along a circle viewed from an eccentric point is exactly the same as a uniform motion along an epicycle combined with a uniform motion of the epicycle can be seen in Fig. 17.3, in which the center of a small circle (the *epicycle*) moves along the larger circle (the *deferent*) in such a way that the angle through which a point on the epicycle has rotated *clockwise* relative to the line joining the center of the epicycle to the center of the deferent equals the angle through which the center of the epicycle has moved *counterclockwise* along the deferent, measured from a fixed diameter of the deferent. Because $OAB'C'$ is a parallelogram, the angle $B'AB$ that an observer at A measures between

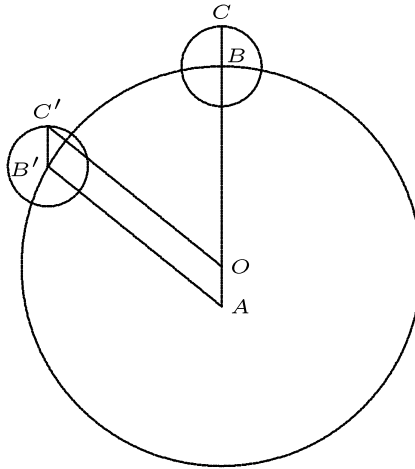


Figure 17.3. Equivalence of epicycles and eccentrics.

the original line AC and the line AB' to the center of the epicycle is exactly the same as the angle that an observer at the center O of the deferent measures between the original line AC and the line to the point C' on the epicycle. Ptolemy demonstrated this equivalence in Section 3 of Book 3. He also considered the possibility that the sun could rotate on its epicycle in the same direction that the epicycle moves around the earth, pointing out that in that case, the most rapid motion would be at apogee (when the sun is farthest from the earth) and the slowest at perigee (when the sun is closest to the earth). In fact, the most rapid motion is at perigee (or perihelion, as we call it, using the heliocentric system). That occurs within a day or two of January 3 each year.⁶

With either model, eccentric or epicyclic with the same angular velocity, no retrograde motion will be observed. The advantage of the epicycle is that retrograde motion can be accounted for by taking the angular velocity on the epicycle greater than that of the epicycle on the deferent. If, for example, the velocity of the planet on the epicycle is twice that of the epicycle on the deferent, retrograde motion will be observed if the radius of the epicycle is more than half the radius of the deferent.

17.2.2. The Motion of the Sun

Since we do not have space to discuss the complexities of the *Almagest*, we shall confine our discussion to a brief sketch of the sun's motion. In this case, the epicycle in Fig. 17.3 can be thought of as showing the approximate positions C and C' of the sun starting in early July (at C) and then at a somewhat later time, around mid-September (at C'). In the eccentric hypothesis, the sun is at B and B' respectively at those times and moving at a uniform rate along the deferent, but the observer (on the earth) is located at A . Whether we imagine an observer at O (the center of the deferent) observing the point C on the epicycle, or an observer at A (displaced from the center of the deferent) observing the point B on the

⁶Perihelion for the center of gravity of the earth–moon system occurs on January 3; but because the two bodies rotate about a common center of gravity, that is not necessarily the day when the center of the earth is at perihelion.

deferent, makes no difference, since both will observe exactly the same amount of rotation at any given time, namely the angle $C'OC$ or the angle $B'AB$.⁷

To summarize: The *mean position* of the sun (B) moves at a constant rate around the deferent, and the deviation from the mean is accounted for either by saying that the earth isn't at the center of the deferent, or that the sun is revolving around its mean position on the epicycle, again at a uniform rate. Either assumption allows the actual motion to be uniform while its appearance to terrestrial observers is not uniform.

The single-epicycle, or eccentric, model is well suited for a comparatively simple motion such as that of the sun. The *path* of the sun among the stars is the ecliptic, which for our purposes is regarded as a fixed circle. Its *motion* along this path, however, is not at a uniform angular rate. It moves most slowly when passing through the constellation Gemini, which astronomers and astrologers refer to as the House of Cancer. (When the houses of the Zodiac were originally established, the House and the constellation of the same name coincided. Because of precession of the equinoxes, they are now about one month out of phase.) The summer solstice in Hellenistic times was in the constellation Cancer, so that the slowest motion of the sun occurred before the solstice, in late May. Nowadays the apogee (aphelion in the heliocentric system) is reached in early July, shortly after the summer solstice (the northernmost point on the ecliptic). Hipparchus placed the apogee about 24° before the summer solstice. Using this information and the fact that (in his day), spring was $94\frac{1}{2}$ days long while summer was $92\frac{1}{2}$ days long, Ptolemy managed to fit the sun's motion by using an epicycle and deferent whose radii were in the ratio of 1 : 24. This ratio, briefly improved upon by Copernicus before Kepler banished circles altogether in favor of ellipses, gave good agreement with observation. Using that scheme and fitting the data to the appropriate dates in 2010, one can obtain the following table of right ascensions of the sun at 30-day intervals throughout the year. The third column of the table gives the values computed by modern astronomy, and the last column shows the amount by which "Ptolemy's" (actually, our) predicted values fall short of the modern values, which is always less than $4'$ of arc, or one degree.⁸

Date	Ptolemy	Modern	Difference
January 19	20 5' 36"	20 6' 53"	1' 17"
February 18	22 7' 54"	22 8' 16"	22"
March 20	00' 4"	00' 4"	0
April 19	1 49' 51"	1 50' 0	9"
May 19	3 44' 54"	3 45' 41"	47"
June 18	5 46' 58"	5 48' 39"	1' 41"
July 18	7 49' 29"	7 52' 1"	2' 31"
August 17	9 45' 0"	9 48' 4"	3' 4"
September 16	11 33' 43"	11 37' 6"	3' 23"
October 16	13 22' 33"	13 26' 7"	3' 34"
November 15	15 20' 2"	15 23' 29"	3' 14"
December 15	17 29' 38"	17 32' 21"	2' 43"

⁷It is impossible to observe the distances to the stars with the unaided eye, so that the only things we can actually measure are the angles between our lines of sight to two different stars. Thus, the distances (radii of the epicycles) can be anything they have to be to account for the angles we actually observe.

⁸A minute in this context is a minute of *time*—one-sixtieth of an hour, and an hour is 15 degrees.

17.3. THE *ALMAGEST*

There is insufficient space here to describe Ptolemy's entire treatise, and in any case our primary concern is with its mathematical innovations. To make geometric astronomy work, Ptolemy developed a subject that resembles what we now call spherical trigonometry, extending earlier work by Hellenistic mathematicians such as Menelaus of Alexandria.

17.3.1. Trigonometry

The word *trigonometry* means *triangle measurement*, but angles are generally measured in terms of the amount of rotation they represent, that is, in terms of the ratio of the length of the arc they subtend to the circumference of the circle containing the arc. That is the context in which Ptolemy developed the subject. It is essentially the study of the quantitative relations between chords and arcs in a circle.

In a system that is still basically the standard one, Ptolemy divides the circumference into 360 equal parts, and measures angles in terms of those parts, that is, in degrees (sometimes half-degrees). The basic problem of trigonometry, from this point of view, is to determine the length of the chord subtended by a given arc and vice versa. To this end, following the Babylonian sexagesimal system, Ptolemy uses $\frac{1}{60}$ of the radius of the circle as the unit of length for chords in a given circle. The effect of this technique is that when two circles intersect, their common chord must be expressed in two different ways, in terms of the two radii. This procedure leads to constant scaling of lengths, and is apt to provoke an impatient reaction from the modern reader. Cumbersome though it was, however, it worked and enabled Ptolemy to give an accurate quantitative description of celestial motions.

17.3.2. Ptolemy's Table of Chords

The computation of the table of chords used by Ptolemy is an interesting exercise in numerical methods. The natural approach would be to start with a central angle whose chord is known (say, 60° , for which the chord equals the radius of the circle), then use half-angle formulas to compute the chord of 30° , 15° , $7^\circ 30'$, and so on, until the desired tabular difference is achieved, after which one would build up the table in these intervals using the addition formulas for the trigonometric functions.⁹ Ptolemy's approach is like this, but he does the computations very elegantly, using what is now called *Ptolemy's theorem*: *In a quadrilateral inscribed in a circle, the rectangle on the diagonals equals the sum of the rectangles on the two pairs of opposite sides*. To prove this theorem, draw a line BE from the vertex B to the diagonal AC such that $\angle ABE = \angle DBC$, as in Fig. 17.4. Hence $\angle EBC = \angle ABD$. Therefore, since angles BAC and BDC are both inscribed in the same arc, triangles ABE and DBC are similar. For the same reason, triangles EBC and ABD are similar. It follows that $AB \cdot CD + BC \cdot AD = AE \cdot BD + EC \cdot BD = AC \cdot BD$.

Ptolemy's theorem makes it possible to express the chord on the difference of two arcs in terms of the chords on the individual arcs. Given three points on a circle, say A , B , and C , take point D diametrically opposite one of the points, say A (see Fig. 17.5). If the chords

⁹In fact the algorithm by which hand calculators evaluate the trigonometric functions works roughly along these lines. Certain values are hard-wired into the calculator and others are computed by application of the addition formulas.

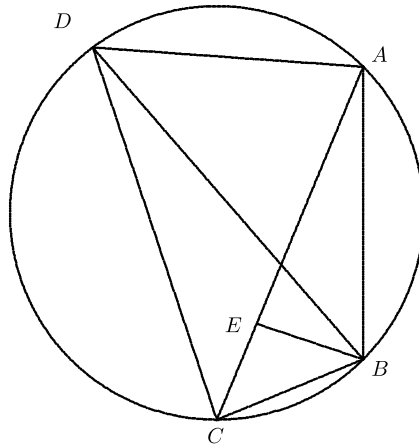


Figure 17.4. Ptolemy's theorem.

AC and AB are given, draw the diameter AD and the chords BC , DB , and CD . The chord AD is known, being the diameter of the circle (hence equal to 120 of Ptolemy's units). Then DB and DC can be computed using the Pythagorean theorem from the diameter and the given chords, since an angle inscribed in a semicircle is a right angle. Hence in the inscribed quadrilateral $ABCD$ both diagonals and all sides except BC are known, and so BC can be computed.

Ptolemy used this theorem to construct a table of chords of angles at half-degree intervals. He began with a regular decagon inscribed in a circle. The central angles subtended by the sides of this decagon are each equal to 36° . Because of the compass-and-straightedge construction of this figure, its side can be expressed as $\frac{\sqrt{5}-1}{2}r$, where r is the radius, which is 60 standard units according to Ptolemy. Instead of repeatedly bisecting this angle, however, Ptolemy adopted an indirect strategy to find the chord of a smaller angle without having to extract so many square roots. He used the fact that the side of the regular pentagon inscribed in a circle (the chord of 72°) is known from Euclid's *Elements*, Book 13, Proposition 10 to be the hypotenuse of the right triangle whose legs are the radius of the circle and the side of the inscribed regular decagon. Thus this chord is

$$2r\sqrt{\frac{5-\sqrt{5}}{8}} = \sqrt{\frac{5-\sqrt{5}}{2}}r,$$

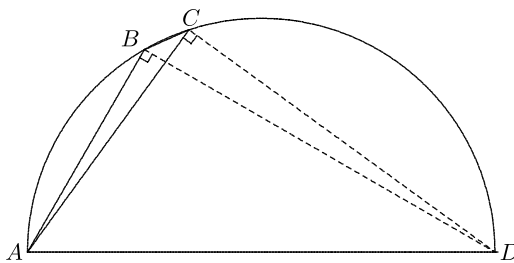


Figure 17.5. The chord of 12° .

which, given that $r = 60$, is 70.5342302751... In Ptolemy's sexagesimal notation, this number is given to the nearest second as 70; 32, 3. Since the chord of 60° is obviously r , which is 60 standard units, one can then use Ptolemy's theorem to compute the chord of $72^\circ - 60^\circ = 12^\circ$. In order to apply this theorem, we first need to get the chords on the arcs of 120° and 108° supplementary to these two angles, as shown in Fig. 17.5. By the Pythagorean theorem, we get

$$\begin{aligned}\text{chord } 120^\circ &= \sqrt{4r^2 - r^2} = \sqrt{3}r; \\ \text{chord } 108^\circ &= \sqrt{4r^2 - \frac{5 - \sqrt{5}}{2}r^2} = \sqrt{\frac{3 + \sqrt{5}}{2}}r.\end{aligned}$$

Thus, the chord of 12° is

$$\left(\sqrt{\frac{15 - 3\sqrt{5}}{8}} - \sqrt{\frac{3 + \sqrt{5}}{8}} \right) r,$$

and given that $r = 60$, this becomes approximately 12.5434155922... Ptolemy gave it as 12; 32, 36.

We have given this computation in language that is more symbolic than Ptolemy's. He always wrote 60 where we have written r , and he had no symbol for the square root. He would first write down the square whose root is to be taken, as a number rather than an expression, and then write down the square root, again as a number.

Ptolemy then showed how to compute the chord of half an angle if the chord of the angle is known. In this way he was able to compute successively the chords of 6° , then 3° , then $1^\circ 30'$, and finally $0^\circ 45'$. He found that, to three sexagesimal places, the chord of $1^\circ 30'$ is 1; 34, 25 and the chord of $0^\circ 45'$ is 0; 47, 8. The ingenious idea of starting from a 72° angle, rather than the more natural 60° angle, allowed Ptolemy to reach angles less than 1° while minimizing the roundoff error caused by approximating square roots.

Ptolemy's construction of his table misses the important angle of 1° . This gap is not accidental. All the angles whose chords can be found by his strategy can be constructed with compass and straightedge, but a 1° angle is not constructible with these instruments alone. To estimate the chord of 1° , Ptolemy combined the two chords on each side of 1° , namely $1^\circ 30'$ and $0^\circ 45'$ with a useful approximation theorem: *The ratio of the larger of two chords to the smaller is less than the ratio of the arcs they subtend.* We have already encountered a theorem similar to this, but not quite identical, in connection with the work of Zenodorus on the isoperimetric problem. (Compare Fig. 16.1 of Chapter 16 with Fig. 17.6 of the present chapter.)

In other words, the ratio of a larger chord to its arc is less than the ratio of a smaller chord to its arc. In still other words, the ratio of chord to arc decreases as the arc increases. In our own language, using radian measure for angles, the chord of an arc of length $r\theta$ is $2r \sin(\theta/2)$. This theorem says that the ratio $\frac{\sin \varphi}{\varphi}$ decreases as φ increases. Because of this proposition, the chord of 1° is smaller than four-thirds of the chord of $0^\circ 45'$, and larger than two-thirds of the chord of $1^\circ 30'$. If we treat Ptolemy's two approximations as exact, we find that the chord of 1° is less than 1; 2, 50, 40 and larger than 1; 2, 50. Ptolemy truncated the first of these to 1; 2, 50. He then wrote (what is logically absurd) that "the chord of 1°

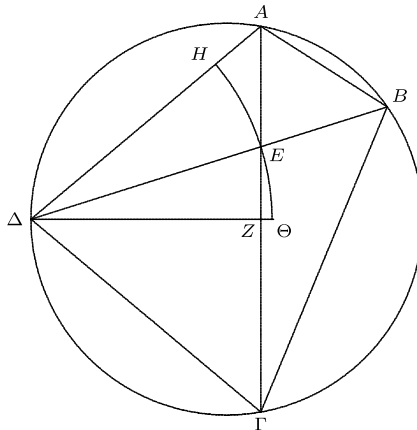


Figure 17.6. A fundamental inequality from Ptolemy’s *Almagest*.

has been shown to be both greater and less than the same amount.” But we know what he means.

Thus Ptolemy had established that the chord of 1° is approximately $1; 2, 50$ units when the radius is 60 units. Then, using his half-angle formula, he found the chord of $0^\circ 30'$ to be $0; 31, 25$, after which he was able to construct the table of chords for angles at half-degree intervals.

The table of chords makes it possible to solve right triangles, in particular, to find the angles in such a triangle when given the ratio of two of its sides. In astronomy, however, one is always using angular coordinates on a sphere, since both the sides and angles of a spherical triangle are given as angles. It would be clumsy always to have to introduce plane triangles in order to find the parts of spherical triangles, and so Ptolemy included certain relations among the parts of spherical triangles as lemmas. These are not the laws of cosines and sines now used in spherical trigonometry, but rather two theorems that had been published half a century earlier in a work called *Sphaerike* by Menelaus of Alexandria. With these relations it is possible to solve such problems as finding which portion of the ecliptic rises simultaneously with a given portion of the celestial equator.

With this mathematical equipment and a wealth of observational data, Ptolemy was able to apply the theoretical methods invented by earlier astronomers. The 12 books of the *Almagest* became the standard astronomical treatise in the Middle East and Europe until the sixteenth century. The details are too complicated to summarize, and we shall have to leave the reader with just the sample given above for the motion of the sun over a single year.

PROBLEMS AND QUESTIONS

Mathematical Problems

- 17.1.** Use Fig. 17.6 to show that the ratio of a larger chord to a smaller is less than the ratio of the arcs they subtend, that is, show that $B\Gamma : AB$ is less than $\widehat{B\Gamma} : \widehat{AB}$, where ΔZ is the perpendicular bisector of $A\Gamma$. (*Hint:* $B\Delta$ bisects angle $AB\Gamma$.) Carry out

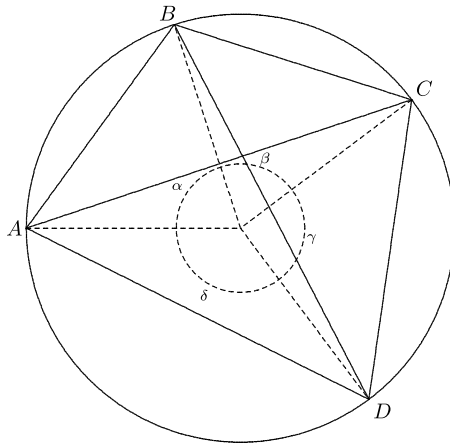


Figure 17.7. Ptolemy's theorem.

the analysis carefully and get accurate upper and lower bounds for the chord of 1° . Convert this result to decimal notation, and compare with the actual chord of 1° which you can find from a calculator. (It is $120 \sin\left(\frac{1^\circ}{2}\right)$.)

- 17.2.** In modern language, the chord of an arc α can be expressed as $d \sin\left(\frac{\alpha}{2}\right)$, where d is the diameter of the circle. Referring to Fig. 17.7, show that Ptolemy's theorem is logically equivalent to the following relation, for any three arcs α , β , and γ of total length less than the full circumference.

$$\sin\left(\frac{\alpha}{2}\right) \cdot \sin\left(\frac{\gamma}{2}\right) + \sin\left(\frac{\beta}{2}\right) \cdot \sin\left(\frac{\alpha + \beta + \gamma}{2}\right) = \sin\left(\frac{\alpha + \beta}{2}\right) \cdot \sin\left(\frac{\beta + \gamma}{2}\right).$$

- 17.3.** Compute the chord of 6° in two different ways: (1) by expressing 6° as the difference of a 36° arc and a 30° arc, whose chords are known to be $30(\sqrt{5} - 1)$ and $30\sqrt{2}(\sqrt{3} - 1)$; (2) by expressing it as the chord of the difference of 12° and 6° .

Historical Questions

- 17.4.** How did the ancients determine geographical latitude?
- 17.5.** Why was geographic longitude more difficult to determine than latitude?
- 17.6.** Out of what mathematical and observational data did Ptolemy construct his astronomical treatise?

Questions for Reflection

- 17.7.** Does the success of Ptolemy's *Almagest* vindicate Plato's conviction that the key to understanding the material world was to connect it with an ideal world of abstractions (ideas or forms) that are perceived with the mind rather than the senses?
- 17.8.** In Ptolemy's system, the occasional retrograde motion of, say Mars, is explained by its motion on the epicycles, whose deferents are at or near the center of the earth. Now,

retrograde motion of the outer planets—westward rather than eastward on (or near) the ecliptic—is observed in the time interval from just before to just after a planet is in opposition to the sun, that is, 180° opposite the sun on the celestial sphere. The fitting of epicycles for any planet must therefore take account of the position of the sun, which itself never undergoes retrograde motion. Surely, one would think, these considerations suggest that the planets are more closely tied to the sun than to the earth, and heliocentric astronomy would be much simpler than geocentric. And in fact, Ptolemy considered this hypothesis, which had been proposed by the astronomer Aristarchus of Samos (ca. 310–ca. 230), but he rejected it on physical grounds. Why did it take another 1500 years for this hypothesis to be revived and become the cornerstone of modern astronomy?

- 17.9.** Ptolemy used a wheel submerged in water up to its axle in order to determine the refraction of light in passing from air into. Based on his observations he gave the following table of the angles of refraction for angles of incidence of 10° , 20° , ..., 80° . (These are the angles the ray makes with the vertical as it enters the water. The angles of refraction are the angles the ray makes with the vertical after entering the water.) The third column gives the values computed from Snell's Law. Since the ratio of the velocities is about 4 : 3 for light in air and light in water, Snell's law says that $\sin(\varphi) = 3 \sin(\theta)/4$, where θ is the angle of incidence and φ is the angle of refraction.

Angle of Incidence	Angle of Refraction	Snell's Law
10°	8°	7.48°
20°	$15\frac{1}{2}^\circ$	14.86°
30°	$22\frac{1}{2}^\circ$	22.02°
40°	29°	28.82°
50°	35°	35.07°
60°	$40\frac{1}{2}^\circ$	40.51°
70°	$45\frac{1}{2}^\circ$	44.81°
80°	50°	47.61°

Since physical scientists, and Ptolemy in particular, are known to have manipulated their data to fit a theory, does this table indicate any such manipulation?

Pappus and the Later Commentators

The last few centuries of mathematics in the Greek tradition showed clear evidence of decline in geometry. Except for Pappus and a few others, most geometric work done during this period was commentary on earlier work. On the other hand, as we saw in Chapter 9, algebra arose in a form that we can recognize, in the works of Diophantus, although it was more closely connected to number theory than to geometry.

18.1. THE COLLECTION OF PAPPUS

Almost nothing is known about the life of Pappus. The tenth-century encyclopedia known as the *Suda* says the following about him:

Pappus, of Alexandria, philosopher, lived about the time of the Emperor Theodosius the Elder [who ruled from 379 to 395], when Theon [of Alexandria] the Philosopher, who wrote the Canon of Ptolemy, also flourished.

However, a table written by Theon himself mentions the emperor Diocletian, who ruled from 284 to 305, and says of his reign, “Pappus wrote during that time.”

The *Suda* is an unreliable source. (We shall see below that it says Hypatia was the wife of Isodoros, who was, like Hypatia, a neo-Platonist philosopher, probably one of the last. He could not have been her husband, since he lived nearly a century later.) Thus, we give a higher credibility to Theon. But there is circumstantial evidence that he also got it wrong, since Pappus wrote a commentary on Ptolemy’s *Almagest* and in it revealed that he had observed an eclipse of the sun on October 18, 320. He is therefore somewhat later than Theon thought and earlier than the author of the *Suda* thought. His most probable dates are from around 290 to 350, and his main work, the *Collection* ($\Sigma\nu\nu\alpha\gamma\omega\gamma\acute{\eta}$ = *Synagōgē*) was written in the early-to-mid fourth century.

Heath describes this work as follows:

Obviously written with the object of reviving the classical Greek geometry, it covers practically the whole field. It is, however, a handbook or guide to Greek geometry rather than an encyclopædia; it was intended, that is, to be read with the original works (where still extant) rather than to enable them to be dispensed with.

In fact, it was too late to revive classical Greek geometry. Its potential had already been exhausted long before, and the only way for geometry to make further progress was to adopt

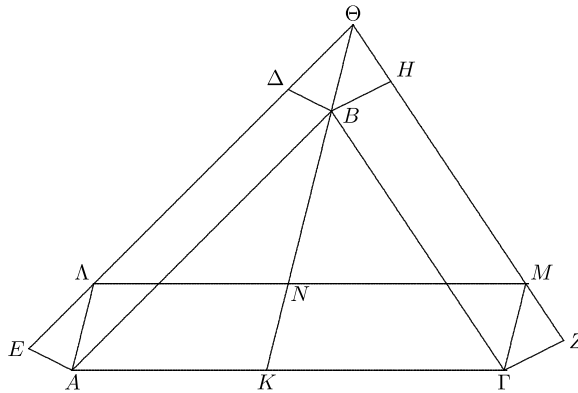


Figure 18.1. Pappus' generalization of the Pythagorean theorem.

new methods. Those methods were not to show up in the West until the time of Fermat, Pascal, and Descartes, some 1300 years later. With that as preface, let us look into the *Collection*.

18.1.1. Generalization of the Pythagorean Theorem

Book 4 of the *Collection* contains a famous generalization of the Pythagorean theorem: Given any triangle $AB\Gamma$ and any parallelograms $B\Gamma ZH$ and $AB\Delta E$ constructed on two sides, it is possible to construct (with straightedge and compass) a parallelogram $A\Gamma M\Delta$ on the third side equal in area to the sum of the other two (see Fig. 18.1).

18.1.2. The Isoperimetric Problem

In Book 5 Pappus states almost verbatim the argument that Theon of Alexandria, quoting Zenodorus, gave for the proof of the isoperimetric inequality. Pappus embroiders the theorem with a beautiful literary device, however. He speaks poetically of the divine mission of the bees to bring from heaven the wonderful nectar known as honey and says that in keeping with this mission they must make their honeycombs without any cracks through which honey could be lost. Being endowed with a divine sense of symmetry as well, the bees had to choose among the regular shapes that could fulfill this condition—that is, triangles, squares, and hexagons. They chose the hexagon because a hexagonal prism required the least material to enclose a given volume, out of all the possible prisms whose base would tile the plane.¹

18.1.3. Analysis, Locus Problems, and Pappus' Theorem

Book 7 of the *Collection* is a treasure trove of fascinating information about Greek geometry for several reasons. First, Pappus describes the kinds of techniques used to carry on the

¹If one is looking for mathematical explanations of this shape, it would be simpler to start with the assumption that the body of a bee is approximately a cylinder, so that the cells should be approximately cylinders. Now one cylinder can be tightly packed with six adjacent cylinders of the same size. If the cylinders are flexible and there is uniform pressure on them, they will flatten into hexagonal prisms.

research that was current at the time. He lists a number of books of this *analysis* and tells who wrote them and what their contents were, in general terms, thereby providing valuable historical information. What he means by analysis, as opposed to synthesis, is a kind of algebraic reasoning in geometry. As he puts it, when a construction is to be made or a relation is to be proved, one imagines the problem to have been solved and then deduces consequences connecting the result with known principles, after which the process is reversed and a proof can be synthesized. This process amounts to thinking about objects determined implicitly in terms of properties that they must have, but not explicitly identified; when applied to numbers—that is, starting with properties that a number must have and deducing its explicit value from those properties—that process is algebra.

A second point of interest in Book 7 is a discussion of locus problems, such as those in Apollonius' *Conics*. This discussion exerted a strong influence on the development of geometry in seventeenth-century France, as we noted in Chapter 15 and will discuss further in Chapter 32. Several propositions from Euclid's *Data*, which was discussed in Section 13.2 of Chapter 13, inspired Pappus to create a very general proposition about plane loci. Referring to the points of intersection of a set of lines, he writes:

To subsume all these discoveries in a single proposition, we have written the following. *If three points are fixed on one line. . . and all the others except one are confined to given lines, then that last one is also confined to a given line.* This is asserted only for four lines, no more than two of which intersect in the same point. It is not known whether this assertion holds for every collection of lines.

This theorem is illustrated in Fig. 18.2, using analytic geometry. To discover this theorem without invoking the power of algebra was an impressive feat. Pappus could not have known that he had provided the essential principle by which a famous theorem of projective geometry known as Desargues' theorem was to be proved 1400 years later. (See Chapter 31.) Desargues knew the work of Pappus, but may not have made the connection with this

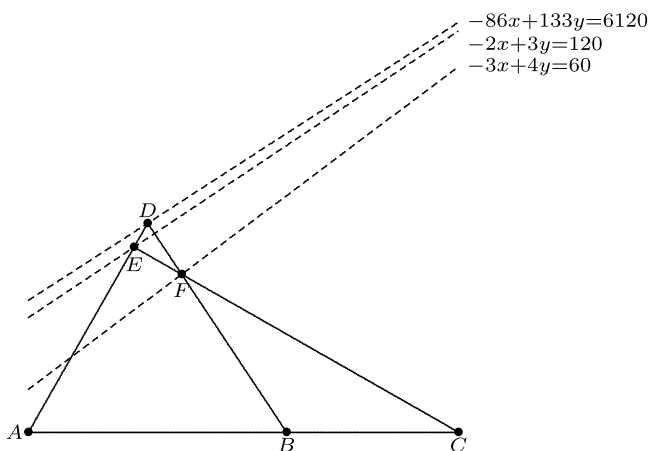


Figure 18.2. Pappus' general locus theorem. The points A , B , and C are fixed at $(0, 0)$, $(90, 0)$, and $(150, 0)$. Point E is confined to the line $-2x + 3y = 120$, F is confined to the line $-3x + 4y = 60$, and E , F , and C are required to be collinear. Then the point D is determined as the intersection of the extensions of the lines AE and BF . The conditions imply that D must lie on the line $-86x + 133y = 6120$.

theorem. The connection was pointed out by van der Waerden (1963, p. 287), who suggests how the theorem may have been proved originally, without analytic geometry.

Pappus discusses the three- and four-line locus for which the mathematical machinery is found in Book 3 of Apollonius' *Conics*. For these cases the locus is always one of the three conic sections. Pappus mentions that the two-line locus is a planar problem; that is, the solution is a line or circle. He says that a point satisfying the conditions of the locus to five or six lines is confined to a definite curve (a curve "given in position" as the Greeks said), but that this curve is "not yet familiar and is merely called a curve." The curve is defined by the condition that the rectangular parallelepiped spanned by the lines drawn from a point to three fixed lines bears a fixed ratio to the corresponding parallelepiped spanned by the lines drawn to three other fixed lines. In our terms, this locus is a cubic curve.

Still a third point of interest is connected with the extension of these locus problems. Pappus considers the locus to more than six lines and says that a point satisfying the corresponding conditions is confined to a definite curve. This step was important, since it proposed the possibility that a curve could be determined by certain conditions without being explicitly constructible. Moreover, it forced Pappus to go beyond the usual geometric interpretation of products of lines as rectangles. Noting that "nothing is subtended by more than three dimensions," he continues:

It is true that some of our recent predecessors have agreed among themselves to interpret such things, but they have not made a meaningful clear definition in saying that what is subtended by certain things is multiplied by the square on one line or the rectangle on others. But these things can be stated and proved using composite ratios.

It appears that Pappus was on the very threshold of the creation of the modern concept of a real number as a ratio of lines. Why did he not cross that threshold? One reason may have been that he was held back by the cumbersome Euclidean definition of a composite ratio, discussed in Section 12.2 of Chapter 12. But there was a further reason: He wasn't interested in foundational questions. He made no attempt to prove or justify the parallel postulate, for example. And that brings us to the fourth attraction of Book 7. In that book Pappus investigated some very interesting problems, which he preferred to foundational questions. After concluding his discussion of the locus problems, he implies that he is merely reporting what other people, who are interested in them, have claimed. "But," he says,

after proving results that are much stronger and promise many applications, . . . to show that I do not come boasting and empty-handed. . . I offer my readers the following: *The ratio of rotated bodies is the composite of the ratio of the areas rotated and the ratio of straight lines drawn similarly [at the same angle] from their centers of gravity to the axes of rotation. And the ratio of incompletely rotated bodies is the composite of the ratio of the areas rotated and the ratio of the arcs described by their centers of gravity.*

The statement of these theorems shows that Pappus is working in the metric-free tradition of Euclid. He does not use the word *volume* at any point, much less say what the volume of any particular solid of revolution is. Instead, he refers only to the ratios of such solids, just as Euclid would have done. To elaborate on this language, if a plane figure A having point P as center of gravity is rotated about a line l , generating a solid X , and a plane figure B having point Q as center of gravity is rotated about a line m , generating a solid Y , and if

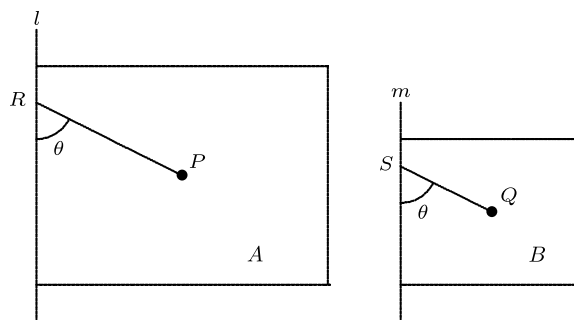


Figure 18.3. Pappus' theorem for two rectangles revolved about an edge. Each of them generates a cylinder whose height equals the side lying on the axis of rotation and whose base has radius equal to the other side. The lines PR and QS are proportional to the perpendicular distances from P and Q to l and m , respectively, so that making them oblique to the axes of rotation does not add any real generality.

line segments PR and QS are drawn from P and Q respectively to points R on l and S on m , making the same angle at R and S , then $X : Y :: A : B.PR : QS$. The simple case when A and B are rectangles and the axes of rotation l and m are edges of A and B respectively is shown in Fig. 18.3.

From Propositions 32 and 34 of Book 11 of the *Elements*, it was known that parallelepipeds having equal bases are proportional to their altitudes and parallelipeds having equal altitudes are proportional to their bases. It is not difficult then to show that the ratio of two parallelepipeds is the composite of the ratios of their bases and altitudes. To do so, let U and V be parallelepipeds having bases A and B , respectively, and altitudes h and k . We wish to show that $U : V :: A : B.h : k$. In accordance with Euclid's definition of the composite ratio, we need three quantities a , b , and c , all of the same kind, such that $A : B :: a : b$ and $h : k :: b : c$. By definition $A : B.h : k$ is $a : c$.

We thus need to show that $U : V :: a : c$ in the notation just introduced. To do so, suppose m and n are any two positive integers such that $ma > nc$. By Archimedes' principle,² there is some positive integer r such that $r(ma - nc) > b$, and hence some integer s such that $rma > sb > rnc$. Since $A : B :: a : b$, it follows that $rmA > sB$; and similarly (because $h : k :: b : c$) we have $sh > rnk$. Therefore the parallelepiped having base rmA and altitude sh is larger than the parallelepiped having base sB and altitude rnk . However, since parallelepipeds having the same base are proportional to their altitudes and vice versa, the former of these is exactly rms times as large as P and the latter is rms times as large as Q . Canceling rs in this statement, we conclude that mP is larger than nQ , which is precisely what the Eudoxan definition of proportion demands. (Of course, the same argument needs to be repeated with the inequality reversed and then repeated again with the inequality replaced by equality in order to satisfy all three of the conditions of the Eudoxan definition of proportion, but that is an easy exercise for the reader.)

Pappus' theorem is easy to prove for rectangles revolved about one of their sides. If the rectangle has sides a and b and is revolved about the side of length a , it generates a cylinder

²This principle says that for any two geometric or physical quantities of the same kind, some integer multiple of each is larger than the other. In modern language: Infinitesimals do not exist. The principle is assumed in Euclid's definition of a ratio.

whose base has radius b and whose altitude is a . It is well known that such a cylinder is proportional to a parallelepiped with square base of side b and altitude a . Now the distance r from the axis to the center of gravity in this case is $\frac{b}{2}$, so that the cylinder is also proportional to a rectangular parallelepiped having a base of sides b and r and altitude a . However, that parallelepiped can also be regarded as having a rectangular base of sides a and b and altitude r . Since we have shown above that parallelepipeds are proportional to the composite ratio of their bases and altitudes, it follows that this cylinder is proportional to the composite of the rectangle of sides a and b and the line r , as asserted. Thus, if we have two such rectangles R_1 and R_2 with sides $a_1, b_1 = 2r_1$ and $a_2, b_2 = 2r_2$, generating cylinders C_1 and C_2 , we can say that $C_1 : C_2 :: R_1 : R_2.r_1 : r_2$. It is then not difficult to get the same theorem for two rectangles, one of which is rotated about any line parallel to one of its sides. You simply “fill in” the space between the rectangle and the axis of rotation with another rectangle so as to make a larger rectangle of the same height. If the original rectangle is $a \times b$ and the filled-in portion is $a \times c$, then the two together will be $a \times (b + c)$. The centroid of the original will be at distance $c + b/2$ from the axis, while the centroid of the one touching the axis of rotation will be at distance $c/2$ from that axis. The centroid of the two together will be at distance $(b + c)/2$ from the axis. Applying the original theorem to the two together, we find that the large cylinder is proportional to a parallelepiped with base $a \times (b + c)$ and altitude $(b + c)/2$, while the cylinder generated by the portion of the rectangle touching the axis of rotation is proportional to a parallelepiped with base that is $a \times c$ and altitude $c/2$. Hence the cylindrical annulus generated by the original rectangle, which is a parallelepiped that has base $a \times b$ and altitude $c + b/2$, is proportional to the difference of these two. The rest of the proof is now easy and is left as an exercise, which the reader may do in modern notation.

What is called Pappus’ theorem in calculus books—and was known for centuries as *Guldin’s formula*—gives a numerical value for the volume generated by revolving a plane region, namely the product of the area and the circumference of the circle traversed by its centroid during the revolution. In this modern form the theorem was first stated in 1609 by the Swiss astronomer/mathematician Paul Guldin (1577–1643), a Jesuit priest, and published between 1635 and 1640 in the second volume of his four-volume work *Centrobaryca seu de centro gravitatis trium specierum quantitatis continuæ* (*The Barycenter, or on the Center of Gravity of the Three Kinds of Continuous Magnitude*). It appears to be established that Guldin had not read Pappus and made the discovery independently. He also gave an inadequate proof of the result, and the first actual proof is due to Bonaventura Cavalieri (1598–1647).

The second result stated by Pappus is an immediate application of the Eudoxan theory of proportion, since the volume generated is obviously in direct proportion to the angle of rotation, as are the arcs traversed by individual points.

In this discussion, we have emphasized that Pappus did not write his results in our modern language of formulas for areas and volumes. Would he have understood them if they had been stated to him? Putting the question another way, how close was he to our point of view? Two concepts that he used now strike us as unnecessary complications. The first was the Euclidean style of avoiding the choice of unit lengths, areas, and volumes. This approach required Pappus to talk about the ratio of two solids rather than the volume of a single solid. He could only say, for example, that solids of revolution are proportional to the composite ratio of the planar regions that are revolved to form them and lines from their centers of gravity to the axes of rotation. That means, to us, that the volume of such a solid is a constant times the product of the area of the planar region and the length of the line, the

same constant for all solids of revolution. In fact, if the angle formed by the line from the center of gravity of the planar region to the axis of rotation is θ , the constant is $2\pi \sin(\theta)$. The other complicated concept is that of a composite ratio. We have now seen two examples of the application of this concept (the argument just given, and Problem 12.3), and it is clear that if two ratios are regarded as numbers, then the composite ratio corresponds to the product of those two numbers. Did Pappus know either of these things? It is very likely that he did, in a sense, although he may not have thought of the situation in quite those terms. As Cuomo (2000, § 5.1) emphasizes, after giving Euclidean-style arguments to prove his propositions, Pappus illustrated many of them with numerical examples. If he had proved this theorem and illustrated it as he did some others, we would have a clearer idea of the extent to which he anticipated the modern refinement of his theorem. But in any case, he was still confined to the notion of a ratio as being a relationship between two objects of the same kind, and he did not think of it as dividing the measure of one of them by the measure of the other. The quantity we obtain by dividing the distance traveled by the time of travel to get the average speed of travel would not have been a ratio to any ancient Greek, and their discussions of motion were not like those of modern physics.

18.2. THE LATER COMMENTATORS: THEON AND HYPATIA

We referred to the later commentators in Chapter 8 as the sources of much of what we know about the history of Greek mathematics. In the present section, we shall say a few words about two of them, namely the fourth-century commentator Theon of Alexandria and his daughter Hypatia, the only woman mathematician of ancient times about whose life a little is known.

18.2.1. Theon of Alexandria

Theon of Alexandria can be dated from the fact that he himself reported that he had observed a solar eclipse on June 16, 364. He continued to write until at least the year 372.

As mentioned above, the tenth-century encyclopedia known as the *Suda* says that Theon lived in the time of the Emperor Theodosius I (379–395). These dates are therefore consistent. It also states that he worked at the Museum at Alexandria (which contained the Library mentioned earlier), as one of its last members apparently, since it did not survive long after his time.

Theon wrote commentaries on many works, including the *Almagest* and the works of Euclid. Until a little over a century ago, his edition of Euclid's *Elements*, on which his daughter Hypatia may have collaborated, was the only known Greek text of the *Elements*. An earlier edition was discovered in the Vatican in the late nineteenth century. From it, historians can see what original contributions were made by Theon and Hypatia. In particular, we see confirmation of what we said earlier: The editors of ancient works were more interested in improving the work than in preserving it intact. O'Connor and Robertson, writing at the MacTutor website,³ relate that Theon elaborated arguments that were obscure in the earlier manuscripts, adding propositions of his own to clarify them, standardized the notation for

³<http://www-history.mcs.st-and.ac.uk/Biographies/Theon.html>

certain concepts, and “corrected errors,” which were not always errors but sometimes mere misunderstandings on the part of Theon.

18.2.2. Hypatia of Alexandria

Very few women mathematicians are known by name from early times. However, Closs (1992, p. 12) mentions a Maya ceramic with a picture of a female scribe/mathematician. From ancient Greece and the Hellenistic culture, two or perhaps three such women are known by name. In his *Lives of Eminent Philosophers*, Diogenes Laertius says that

Pythagoras had a wife named Theano. She was the daughter of Brontinus of Croton, although some say that she was Brontinus’ wife and Pythagoras’ pupil. He also had a daughter named Damo, as Lysis mentions in a letter to Hipparchus. In this letter he speaks of Pythagoras as follows: ‘And many say that you [Hipparchus] give public lectures on philosophy, as Pythagoras once did. He entrusted his *Commentaries* to Damo, his daughter, and told her not divulge them to anyone not of their household. And she refused to part with them, even though she could have sold them for a considerable amount of money. . . .’

Since it was said that the Pythagoreans admitted women to their councils, it seems possible that Pythagoras’ wife and daughter may have engaged in mathematical research. However, nothing at all is known about any works they may have produced. All that we know about them is contained in the paragraph from Diogenes Laertius just quoted.

There are two primary sources for information about the life of Hypatia. One is a passage in a seven-book history of the Christian Church written by Socrates Scholasticus, who was a contemporary of Hypatia but lived in Constantinople; the other is an article in the *Suda*. In addition, several letters of Synesius, bishop of Ptolemais (in what is now Libya), who was a student of Hypatia, were written to her or mention her, always in terms of high respect. In one letter he requests her, being in the “big city,” to procure him a scientific instrument (hygrometer) not available in the less urbanized area where he lived. In another he asks her judgment on whether to publish two books that he had written, saying

If you decree that I ought to publish my book, I will dedicate it to orators and philosophers together. The first it will please, and to the other it will be useful, provided of course that it is not rejected by you, who are really able to pass judgment. If it does not seem to you worthy of Greek ears, if, like Aristotle, you prize truth more than friendship, a close and profound darkness will overshadow it, and mankind will never hear it mentioned. [Fitzgerald, 1926]

The account of Hypatia’s life written by Socrates Scholasticus occupies Chapter 15 of Book 7 of his *Ecclesiastical History*. Socrates Scholasticus describes Hypatia as the pre-eminent philosopher of Alexandria in her own time and a pillar of Alexandrian society, who entertained the elite of the city in her home. Among that elite was the Roman procurator Orestes. There was conflict at the time among Christians, Jews, and pagans in Alexandria; Cyril, the bishop of Alexandria, was apparently in conflict with Orestes. According to Socrates, a rumor was spread that Hypatia prevented Orestes from being reconciled with Cyril. This rumor caused some of the more volatile members of the Christian community to seize Hypatia and murder her in March of 415.

The *Suda* devotes a long article to Hypatia, repeating in essence what was related by Socrates Scholasticus. It says, however, that Hypatia was the wife of the philosopher Isodoros, which is definitely not the case, since Isodoros lived near the end of the fifth century. (He was perhaps the last neo-Platonist in Alexandria.) The *Suda* assigns the blame for her death to Cyril himself.

Yet another eight centuries passed, and Edward Gibbon came to write the story in his *Decline and Fall of the Roman Empire* (Chapter XLVII). In Gibbon's version, Cyril's responsibility for the death of Hypatia is reported as fact, and the murder itself is described with certain gory details for which there is no factual basis.

As for her mathematical works, we have already mentioned that she may have been the editor of some of the books of *Arithmetica* written by Diophantus. From other commentators, it is known that, in addition to her lectures on philosophy, she wrote commentaries on the works of earlier mathematicians.

A fictionalized version of Hypatia's life can be found in a nineteenth-century novel by Charles Kingsley, bearing the title *Hypatia, or New Foes with an Old Face*. What facts are known were organized into an article by Michael Deakin (1994) and a study of her life by Maria Dzielska (1995).

PROBLEMS AND QUESTIONS

Mathematical Problems

- 18.1.** Prove Pappus' generalization of the Pythagorean theorem, shown in Fig. 18.1, assuming any parallelogram $AB\Delta E$ whatsoever and any parallelogram $B\Gamma ZH$ whatsoever have been constructed on sides AB and $B\Gamma$. Extend the outer sides of these two parallelograms to Θ , draw the line ΘB , and extend it to meet $A\Gamma$ at K . Draw ΓM and $A\Lambda$ parallel to ΘK , meeting ZH and $E\Delta$ in M and Λ respectively. It is easy to see that $\Theta M\Gamma B$ and $\Theta\Lambda AB$ are parallelograms and that therefore $A\Lambda = B\Theta = \Gamma M$. Hence if ΛM is drawn, meeting BK in N , we shall also have $KN = B\Theta$. Then prove that $A\Gamma M\Lambda = AB\Delta E + B\Gamma ZH$.
- 18.2.** Explain why Pappus' generalization of the Pythagorean theorem is not merely a trivial consequence of Proposition 31 of Book 6 of the *Elements*, which states that any similar polygons, similarly situated on the three sides of a right triangle, satisfy the same relation as squares; that is, the sum of the two figures on the legs equals the figure on the hypotenuse.
- 18.3.** Prove Guldin's formula for the union of two regions, given that it is true for each of them. If the centroid of an area A lies at distance r from the axis, and the centroid of another area B disjoint from it lies at distance s from that axis (both areas being on the same side of the axis), then the centroid of the union of the two areas lies at distance $\rho = \frac{Ar+B s}{A+B}$ from the axis. (By Archimedes' principle of the lever, this is the distance from the axis to the point at which weights proportional to A and B in the given locations will balance, since it differs from r by $|r - \rho| = \frac{B}{A+B}|r - s|$ and from s by $|s - \rho| = \frac{A}{A+B}|r - s|$, distances that are inversely proportional to A and B .) For any plane area that can be approximated from within and without by a union of rectangles, the method of exhaustion then yields the Guldin formula.

Historical Questions

- 18.4.** Describe some mathematical results that are found in Pappus' *Collection*.
- 18.5.** For what contributions to mathematics is Theon of Alexandria remembered?
- 18.6.** What position in Alexandrian society did Hypatia have?

Questions for Reflection

- 18.7.** By no means all the conceivable theorems of metric-free plane geometry are found in the works of the authors we have discussed. One that arose in the nineteenth century—known as the Steiner–Lehmus theorem after Jacob Steiner (1796–1863), who proved it, and D. C. L. Lehmus (1780–1863), who posed it—asserts that if two angle bisectors of a triangle are equal, then the triangle is isosceles.⁴ There are many such results, including, for example, the 1899 discovery by Frank Morley (1860–1937) that the trisectors of the angles of a triangle intersect inside the triangle in three points that are the vertices of an equilateral triangle. Do such results mean that in fact Greek geometry didn't decline at all, that it is still alive and well?
- 18.8.** Sketch out a historical-fiction scenario in which the ancient mathematicians discover analytic geometry through the locus problems discussed by Pappus. To do so, they would have to learn how to interpret a ratio of lines as what we call a real number. Explain how they could have interpreted multiplication and division of such ratios.
- 18.9.** What is meant by saying that Greek geometry was in decline after the time of Apollonius? What did this decline amount to, and how could geometry have been revived?

⁴This theorem makes a very nice puzzle for the amateur geometer. High-school students sometimes produce very clever proofs of it, but it is deceptively difficult to prove.

INDIA, CHINA, AND JAPAN 500 BCE–1700 CE

In the six chapters that constitute this part, we shall survey a long period of development of mathematics in three cultures that grew up independent of the mathematics that flourished around the Mediterranean Sea. A different “flavor,” more numerically oriented and strongly algebraic, will be seen in all three places. This numerical orientation will be especially noted in geometry, where the approach is not the axiomatic, metric-free Euclidean system. Nonobvious relations among geometric figures are demonstrated using congruence, dissection, and the Pythagorean theorem.

The usual disclaimer applies in this part. It is a very small sample of what could be said; and for more details, the reader should consult the references cited in the corresponding chapters.

Contents of Part IV

1. Chapter 19 (Overview of Mathematics in India) contains a survey of some major achievements and outstanding mathematicians in India (including modern Pakistan) from the earliest times to the twentieth century. By looking at the prefaces to some of the great treatises, we gain some idea of the motivation for creating this knowledge. Again, the names mentioned are only a few of a large number that are worthy of mention.
2. Chapter 20 (From the *Vedas* to Aryabhata I) discusses the mathematics of the Hindu *Vedas* beginning around 500 BCE and the work of Aryabhata I (476–550).
3. Chapter 21 (Brahmagupta, the *Kuttaka*, and Bhaskara II) discusses the work of two more outstanding mathematicians from the seventh through twelfth centuries CE: Brahmagupta (598–670), and Bhaskara II (1114–1184). We end the story at that point, even though mathematics continued to flourish in India with no break at the end of the twelfth century, even anticipating some parts of the calculus.
4. Chapter 22 (Survey of Chinese Mathematics) is devoted to the Chinese development of arithmetic, algebra, and geometry to meet practical administrative needs. The treatises involved include the ancient *Zhou Bi Suan Jing* (*Arithmetical Classic of the Zhou*), which probably dates from a time earlier than 200 BCE, and the Han-Dynasty (200 BCE–200 CE) document *Jiu Zhang Suanshu* (*Nine-Chapter Mathematical Treatise*), which can be regarded as the fundamental text on classical Chinese mathematics.

5. Chapter 23 (Later Chinese Algebra and Geometry) discusses the study of higher-order equations by Chinese mathematicians and the advanced geometry of Liu Hui (220–280), Zu Chongzhi (420–501), and Zu Geng (ca. 450–ca. 520).
6. Chapter 24 (Traditional Japanese Mathematics) discusses the mathematics of Japan as it was developed from the Chinese classic works and elaborated during the Tokugawa Era from 1600 through 1867. This subject, called *wasan* (Japanese-style computation), is contemporaneous with a phenomenon that is apparently unique to Japan, namely, the hanging of votive plaques at Shinto and Buddhist shrines with worked-out mathematical problems on them. These plaques are called *sangaku* (computational framed pictures).

Overview of Mathematics in India

From archaeological excavations at Mohenjo Daro and Harappa on the Indus River in Pakistan it is known that an early civilization existed in this region for about a millennium starting in 2500 BCE. This civilization may have been an amalgam of several different cultures, since anthropologists recognize five different physical types among the human remains. Many of the artifacts that were produced by this culture have been found in Mesopotamia, evidence of trade between the two civilizations. As a framework for the mathematical history we shall be studying in this chapter and the two following, we shall periodize this history as follows.

1. *The Aryan Civilization.* The early civilization of these five groups of people disappeared around 1500 BCE, and its existence was not known in the modern world until 1925. The cause of its extinction is believed to be an invasion from the northwest by a sixth group of people, who spoke a language closely akin to early Greek. Because of their language, these people are referred to as Aryans, a term that acquired a sinister racial meaning in the early twentieth century. (Used in this sense, it was a blemish on a popular brief history of mathematics by W. W. Rouse Ball.) The Aryans gradually expanded and formed a civilization of small kingdoms, which lasted about a millennium.
2. *Sanskrit Literature.* The language of the Aryans became a literary language known as Sanskrit, in which great classics of literature and science have been written. Sanskrit thus played a role in southern Asia analogous to that of Greek in the Mediterranean world and Chinese in much of eastern Asia.¹ That is, it provided a means of communication among scholars whose native languages were not mutually comprehensible and a basis for a common literature in which cultural values could be preserved and transmitted. During the millennium of Aryan dominance, the spoken language of the people gradually diverged from written Sanskrit. Modern descendants of Sanskrit are Hindi, Gujarati, Bengali, and others. Sanskrit is the language of the *Mahabharata* and the *Ramayana*, two epic poems whose themes bear some resemblance to the

¹India also exerted a huge cultural influence on southeast Asia, through the Buddhist and Hindu religions and in architecture and science. Moreover, both cultural contact and commercial contact between India and China have a long history.

Homeric epics, and of the *Upanishads*, which contain much of the moral teaching of Hinduism.

Among the most ancient works of literature in the world are the Hindu *Vedas*. The word means *knowledge* and is related to the English word *wit*. The composition of the *Vedas* began around 900 BCE, and additions continued to be made to them for several centuries. Some of these *Vedas* contain information about mathematics.

3. *Hindu Religious Reformers*. Near the end of the Aryan civilization, in the second half of the sixth century BCE, two figures of historical importance arise. One of these was Gautama Buddha (563–479), heir to a kingdom near the Himalaya Mountains, whose spiritual journey through life led to the principles of Buddhism. The other, named Mahavira (599–527), is less well known but has some importance for the history of mathematics. Like his contemporary Buddha, he began a reform movement within Hinduism. This movement, known as Jainism, still has several million adherents in India. It is based on a metaphysic that takes very seriously what is known in some Western ethical systems as the *chain of being*. Living creatures are ranked according to their awareness. Those having five senses are the highest, and those having only one sense are the lowest.
4. *Islam in India*. The rapid Muslim expansion from the Arabian desert in the seventh century brought Muslim invaders to India by the early eighth century. The southern valley of the Indus River became a province of the huge Umayyad Empire, but the rest of India preserved its independence, as it did 300 years later when another Muslim people, the Turks and Afghans, invaded. Still, the contact was enough to bring certain Hindu works, including the Hindu numerals, to the great center of Muslim culture in Baghdad. The complete and destructive conquest of India by the Muslims under Timur the Lame came at the end of the fourteenth century. Timur did not remain in India but sought new conquests; eventually he was defeated by the Ming dynasty in China. India was desolated by his attack and was conquered a century later by Akbar the Lion, the first of the Mogul emperors and a descendant of both Genghis Khan and Timur the Lame. The Mogul Empire lasted nearly three centuries and was a time of prosperity and cultural resurgence. One positive effect of this second Muslim expansion was a further exchange of knowledge between the Hindu and Muslim worlds. Interestingly, the official administrative language used for Muslim India was neither Arabic nor an Indian language; it was Persian.
5. *British Rule*. During the seventeenth and eighteenth centuries British and French trading companies were in competition for the lucrative trade with the Mogul Empire. British victories during the Seven Years War (1756–1763) left Britain in complete control of this trade. Coming at the time of Mogul decline due to internal strife among the Muslims and continued resistance on the part of the Hindus, this trade opened the door for the British to make India part of their empire. British colonial rule lasted nearly 200 years, coming to an end only after World War II. British rule made it possible for European scholars to become acquainted with Hindu classics of literature and science. Many Sanskrit works were translated into English in the early nineteenth century and became part of the world's science and literature.
6. *Independent India*. Some 65 years have now passed since India became an independent nation. This period has been one of great cultural and economic resurgence in India, and mathematics has benefited fully from this resurgence.

Within this general framework, we can distinguish three periods in the development of mathematics in the Indian subcontinent. The first period begins around 900 BCE with individual mathematical results forming part of the *Vedas*. The second begins with treatises called *siddhantas*, concerned mostly with astronomy but containing explanations of mathematical results, which appear in the second century CE. These treatises led to continuous progress for 1500 years, during which time much of algebra, trigonometry, and certain infinite series that now form part of calculus were discovered, a century or more before Europeans developed calculus. In the third stage, which began during the two centuries of British rule, this Hindu mathematics came to be known in the West, and Indian mathematicians began to work and write in the modern style of mathematics that is now universal. In the present chapter, we shall discuss this mathematical development in general terms, concentrating on a few of the major works and authors and their motivation, with mathematical details to follow in the succeeding chapters.

19.1. THE *SULVA SUTRAS*

In the period from 800 to 500 BCE a set of verses of geometric and arithmetic content were written and became part of the *Vedas*. These verses are known as the *Sulva Sutras* or *Sulba Sutras*.² The name means *Cord Rules* and probably reflects the use of a stretched rope or cord as a way of measuring length, as in Egypt. The root *sulv* originally meant *to measure* or *to rule*, although it also has the meaning of a cord or rope; *sutra* means *thread* or *cord*, a common measuring instrument. In the case of the *Vedas* the objects being measured with the cords were altars. The maintenance of altar fires was a duty for pious Hindus; and because Hinduism is polytheistic, it was necessary to consider how elaborate and large the fire dedicated to each deity was to be. This religious problem led to some interesting problems in arithmetic and geometry.

Two scholars who studied primarily the Sanskrit language and literature made important contributions to mathematics. Pingala, who lived around 200 BCE, wrote a treatise known as the *Chandahsutra*, containing one very important mathematical result, which, however, was stated so cryptically that one must rely on a commentary written 1200 years later to know what it meant. Later, a fifth-century scholar named Panini standardized the Sanskrit language, burdening it with some 4000 grammatical rules that make it many times more difficult to learn than any other Indo-European language. In the course of doing so, he made extensive use of combinatorics and the kind of abstract reasoning that we associate with algebra. These subjects set the most ancient Hindu mathematics apart from that of other nations.

²The *Sulva Sutras* are discussed in many places. The reader is cautioned against books discussing “Vedic mathematics,” however—for example, Maharaja (1965), which presents elaborate modern mathematical arguments tenuously connected to the *Vedas* and alleges that analytic geometry in its modern form, which associates an equation with a curve, was known to the Vedic authors 2500 years ago. Communication problems occur even in generally reliable sources, such as the book of Srinivasiengar (1967), which is the source of many of the facts reported in this chapter and the next. (Everything in these two chapters comes from some secondary source, usually the books of Srinivasiengar, Plofker, Colebrooke, or Clark.) Srinivasiengar asserts (p. 6) that the unit of length known as the *vyayam* was “about 96 inches,” and “possibly this represented the height of the average man in those days.” This highly improbable statement results from the imprecision in the term *height*. According to Plofker (2009, p. 18), there was a unit called *man height*, but it meant the height a man could reach into the air standing on the ground. Even with that clarification, 96 inches seems improbable.

19.2. BUDDHIST AND JAIN MATHEMATICS

As with any religion that encourages quiet contemplation and the renunciation of sensual pleasure, Jainism often leads its followers to study mathematics, which provides a different kind of pleasure, one appealing to the mind. There have always been some mathematicians among the followers of Jainism, right down to modern times, including one in the ninth century bearing the same name as the founder of Jainism. This other Mahavira speculated on arithmetic operations that yield infinite or infinitesimal results, a topic of interest to Jains in connection with cosmology and physics (Plofker 2009, pp. 58, 163). The early work of Jain mathematicians is notable for algebra (the *Sthananga Sutra*, from the second century BCE), for its concentration on topics that are unique to early Hindu mathematics, such as combinatorics (the *Bhagabati Sutra*, from around 300 BCE), and for speculation on infinite numbers (the *Anuyoga Dwara Sutra*, probably from the first century BCE). The Jains were the first to use the square root of 10 as an approximation to the ratio of a circle to its diameter, that is, the number we call π (Plofker 2009, p. 59). Like the Jains, Buddhist monks were very fond of large numbers, and their influence was felt when Buddhism spread to China in the sixth century CE.

19.3. THE BAKSHALI MANUSCRIPT

A birchbark manuscript unearthed in 1881 in the village of Bakshali, near Peshawar, Pakistan is believed by some scholars to date from the seventh century CE, although Sarkor (1982) believes it cannot be later than the end of the third century, since it refers to coins named *dīnāra* and *dramma*, which are undoubtedly references to the Greek coins known as the denarius and the drachma, introduced into India by Alexander the Great. These coins had disappeared from use in India by the end of the third century. Plofker, however (2009, p. 157) places it somewhere in the period 700–1200. The Bakshali manuscript contains some interesting algebra that will be discussed in Chapter 20.

19.4. THE *SIDDHANTAS*

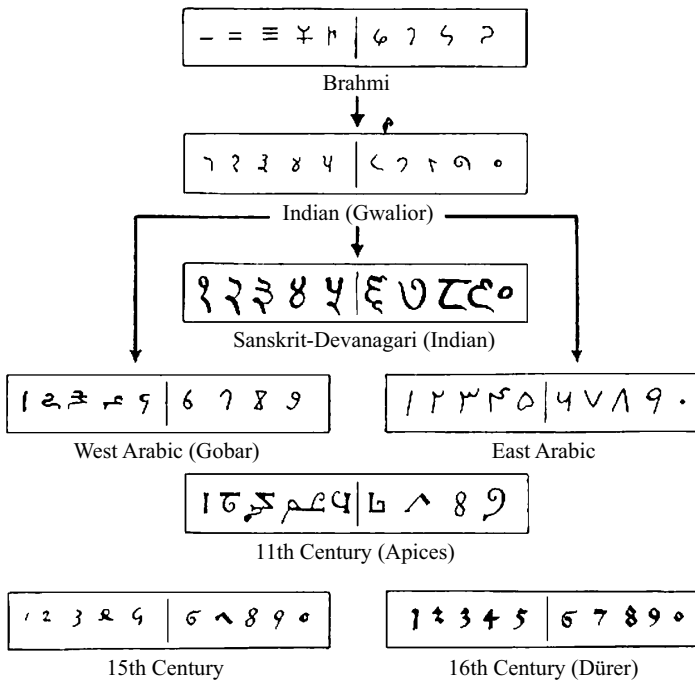
During the second, third, and fourth centuries CE, Hindu scientists compiled treatises on astronomy known as *siddhantas*. The word *siddhanta* means a system.³ One of these treatises, the *Surya Siddhanta* (*System of the Sun*), from the late fourth century, has survived intact. Another from approximately the same time, the *Paulisha Siddhanta*, was frequently referred to by the Muslim scholar al-Biruni (973–1048). The name of this treatise seems to have been bestowed by al-Biruni, who says that the treatise was written by an Alexandrian astrologer named Paul.

19.5. HINDU–ARABIC NUMERALS

The decimal system of numeration, in which 10 symbols are used and the value of a symbol depends on its physical location relative to the other symbols in the representation of a

³A colleague of the author suggested that this word may be cognate with the Greek *idōn* (neuter plural *idōnta*), the aorist participle of the verb meaning *see*, which can be translated as “after seeing . . .”.

number, came to the modern world from India by way of the medieval Muslim civilization. These symbols have undergone some changes in their migration from ancient India to the modern world, as shown in the photo. The idea of using a symbol for an empty place was the final capstone on the creation of a system of counting and calculation that is in all essential aspects the one still in use. This step must have been taken well over 1500 years ago in India. There is some evidence, not conclusive, that symbols for an empty place were used earlier, but no such symbol occurs in the work of Arbyabhata I. On the other hand, such a symbol, called in Sanskrit *sunya* (empty), occurs in the work of Brahmagupta a century after Arybhata.



Evolution of the Hindu–Arabic numerals from India to modern Europe. Copyright © Vandenhoeck & Ruprecht, from the book by Karl Menninger, *Zahlwort und Ziffer*, 3rd ed., Göttingen, 1979.

19.6. ARYABHATA I

With the writing of treatises on mathematics and astronomy, we at last come to some records of the motives that led people to create Hindu mathematics, or at least to write expositions of it. A mathematician named Aryabhata (476–550), the first of two mathematicians bearing that name, lived in the late fifth and early sixth centuries at Kusumapura (now Pataliputra, a village near the city of Patna) and wrote a book called the *Aryabhatiya*. This work had been lost for centuries when it was recovered by the Indian scholar Bhau Daji (1822–1874) in 1864. Scholars had known of its existence through the writings of commentators and had

been looking for it. Writing in 1817, the English scholar Henry Thomas Colebrooke (1765–1837), who translated other Sanskrit mathematical works into English, reported, “A long and diligent research of various parts of India has, however, failed of recovering any part of the. . . *Algebra* and other works of Aryabhata.” Ten years after its discovery the *Aryabhatiya* was published at Leyden and attracted the interest of European and American scholars. It consists of 123 stanzas of verse, divided into four sections, of which the first, third, and fourth are concerned with astronomy and the measurement of time.

Like all mathematicians, Aryabhata I was motivated by intellectual interest. This interest, however, was closely connected with his Hindu piety. He begins the *Aryabhatiya* with the following tribute to the Hindu deity:

Having paid reverence to Brahman, who is one but many, the true deity, the Supreme Spirit, Aryabhata sets forth three things: mathematics, the reckoning of time, and the sphere. [Clark, 1930, p. 1]

The translator adds phrases to explain that Brahman is one as the sole creator of the universe, but is many via a multitude of manifestations.

Aryabhata then continues his introduction with a list of the astronomical observations that he will be accounting for and concludes with a promise of the reward awaiting the one who learns what he has to teach:

Whoever knows this *Dasagitika Sutra* which describes the movements of the earth and the planets in the sphere of the asterisms passes through the paths of the planets and asterisms and goes to the higher Brahman. [Clark, 1930, p. 20]

As one can see, students in Aryabhata’s culture had an extra reason to study mathematics and astronomy, beyond the concerns of practical life and the pleasures of intellectual edification. Learning mathematics and astronomy helped to advance the soul through the cycle of reincarnations that Hindus believed in.

After setting out his teaching on the three subjects, Aryabhata concludes with a final word of praise for the Hindu deity and invokes divine endorsement of his labors:

By the grace of God the precious sunken jewel of true knowledge has been rescued by me, by means of the boat of my own knowledge, from the ocean which consists of true and false knowledge. He who disparages this universally true science of astronomy, which formerly was revealed by Svayambhu⁴ and is now described by me in this *Aryabhatiya*, loses his good deeds and his long life. [Clark, 1930, p. 81]

19.7. BRAHMAGUPTA

The establishment of research centers for astronomy and mathematics at Kusumapura and Ujjain, near the geographical center of modern India, produced a succession of good mathematicians and mathematical works for many centuries after Aryabhata I. Half a century after the death of Aryabhata I, another Hindu mathematician, Brahmagupta (598–670), was born

⁴According to the *Matsya Purana*, the sixteenth purana of the Hindu scriptures, Svayambhu was a self-generated deity who infused the universe with the potential to generate life.

in the city of Sind, now in Pakistan. He was primarily an astronomer, but his astronomical treatise, the *Brahmasphutasiddhanta* (literally *The Corrected Brahma Siddhanta*), contains several chapters on computation (*ganita*). The Hindu interest in astronomy and mathematics continued unbroken for several centuries, producing important work on trigonometry in the tenth century.

19.8. BHASKARA II

Approximately 500 years after Brahmagupta, in the twelfth century, the mathematician Bhaskara (1114–1185), the second of that name, was born on the site of the modern city of Bijapur, in southwestern India. He is the author of the *Siddhanta Siromani*, in four parts, a treatise on algebra and geometric astronomy. Only the first of these parts, known as the *Lilavati*, and the second, known as the *Vija Ganita*,⁵ concern us here. Bhaskara says that his work is a compendium of knowledge, a sort of textbook of astronomy and mathematics. The name *Lilavati* was common among Hindu women. Many of the problems are written in the form of puzzles addressed to this Lilavati.

Bhaskara II apparently wrote the *Lilavati* as a textbook to form part of what we would call a liberal education. His introduction reads as follows:

Having bowed to the deity, whose head is like an elephant's [Ganesh], whose feet are adored by gods; who, when called to mind, relieves his votaries from embarrassment; and bestows happiness on his worshippers; I propound this easy process of computation, delightful by its elegance, perspicuous with words concise, soft and correct, and pleasing to the learned. [Colebrooke, 1817, p. 1]

As a final advertisement at the end of his book, Bhaskara extols the pleasure to be derived from learning its contents:

Joy and happiness is indeed ever increasing in this world for those who have *Lilavati* clasped to their throats, decorated as the members are with neat reduction of fractions, multiplication, and involution, pure and perfect as are the solutions, and tasteful as is the speech which is exemplified. [Colebrooke, 1817, p. 127]

The *Vija Ganita* consists of nine chapters, in the last of which Bhaskara tells something about himself and his motivation for writing the book:

On earth was one named Maheswara, who followed the eminent path of a holy teacher among the learned. His son Bhaskara, having from him derived the bud of knowledge, has composed this brief treatise of elemental computation. As the treatises of algebra [*vija ganita*] by Brahmagupta, Shidhara and Padmanabha are too diffusive, he has compressed the substance of them in a well-reasoned compendium for the gratification of learners. . . to augment wisdom and strengthen confidence. Read, do read, mathematician, this abridgement, elegant in style, easily understood

⁵This Sanskrit word means literally *seed computation*, the word *seed* being used in the algebraic sense of *root*. It is compounded from the Sanskrit root *vij-* or *bij-*, which means *seed*. As we have stated many times, the basic idea of algebra is to name explicitly one or more numbers (the "seed") given certain implicit descriptions of them (metaphorically, "flowers" that they produce), usually the result of operating on them in various ways. The word is usually translated as *algebra*.

by youth, comprising the whole essence of computation, and containing the demonstration of its principles, replete with excellence and void of defect. [Colebrooke, 1817, pp. 275–276]

The mathematician “Shidhara” is probably Sridhara (870–930). Information on a mathematician named Padmanabha does not appear to be available.

19.9. MUSLIM INDIA

Indian mathematical culture reflects the religious division between the Muslim and Hindu communities to some extent. The Muslim conquest brought Arabic and Persian books on mathematics to India. Some of these works were translated from ancient Greek, and among them was Euclid’s *Elements*. These translations of later editions of Euclid contained certain obscurities and became the subject of commentaries by Indian scholars. Akbar the Lion decreed a school curriculum for Muslims that included three-fourths of what was known in the West as the quadrivium. Akbar’s curriculum included arithmetic, geometry, and astronomy, leaving out only music.⁶ Details of this Indian Euclidean tradition are given in the paper by De Young (1995).

19.10. INDIAN MATHEMATICS IN THE COLONIAL PERIOD AND AFTER

One of the first effects of British rule in India was to acquaint European scholars with the treasures of Hindu mathematics described above. A century passed before the British colonial rulers began to establish European-style universities in India. According to Varadarajan (1983), these universities were aimed at producing government officials, not scholars. As a result, one of the greatest mathematical geniuses of all time, Srinivasa Ramanujan (1887–1920), was not appreciated and had to appeal to mathematicians in Britain to gain a position that would allow him to develop his talent. The necessary conditions for producing great mathematics were present in abundance, however, and the establishment of the Tata Institute in Bombay (now Mumbai) and the Indian Statistical Institute in Calcutta were important steps in this direction. After Indian independence was achieved, the first prime minister, Jawaharlal Nehru (1889–1964), made it a goal to achieve prominence in science. This effort has been successful in many areas, including mathematics. The names of Komaravolu Chandrasekharan (b. 1920), Harish-Chandra (1923–1983), and others have become celebrated the world over for their contributions to widely diverse areas of mathematics.

19.10.1. Srinivasa Ramanujan

The topic of power series is one in which Indian mathematicians had anticipated some of the discoveries in seventeenth- and eighteenth-century Europe. It was a facility with this technique that distinguished Ramanujan, who taught himself mathematics after having been refused admission to universities in India. After publishing a few papers, starting in 1911,

⁶The quadrivium is said to have been proposed by Archytas, who apparently communicated it to Plato when the latter was in Sicily to consult with the ruler of Syracuse; Plato incorporated it in his writings on education, as discussed in Chapter 12.

he was able to obtain a stipend to study at the University of Madras. In 1913 he took the bold step of communicating some of his results to G. H. Hardy (1877–1947). Hardy was so impressed by Ramanujan’s ability that he arranged for Ramanujan to come to England. Thus began a collaboration that resulted in seven joint papers with Hardy, while Ramanujan alone was the author of some 30 others. He rediscovered many important formulas and made many conjectures about functions such as the hypergeometric function that are represented by power series.

Unfortunately, Ramanujan was in frail health, and the English climate did not agree with him. Nor was it easy for him to maintain his devout Hindu practices so far from his normal Indian diet. He returned to India in 1919, but succumbed to illness the following year. Ramanujan’s notebooks have been a subject of continuing interest to mathematicians. Hardy passed them on to G. N. Watson (1886–1965), who published a number of “theorems stated by Ramanujan.” The full set of notebooks was published in the mid-1980s (see Berndt, 1985).

QUESTIONS

Historical Questions

- 19.1. What application motivates the mathematics included in the *Sulva Sutras*?
- 19.2. What mathematical subjects studied by Indian mathematicians long ago have no counterpart in the other cultures studied up to this point?
- 19.3. Which physical science is most closely connected with mathematics in the Hindu documents?
- 19.4. What justifications for the study of mathematics do the Hindu authors Aryabhata I and Bhaskara II mention?

Questions for Reflection

- 19.5. What differences do you notice in the “style” of mathematics in Greece and India? Consider in particular the importance of logic, the metaphysical views of the nature of such things as lines, circles, and the like, and the interpretation of the infinite.
- 19.6. One reflection of Mesopotamian influence in India is the division of the circle into 360 degrees. Does having this system in common indicate that the Hindus received their knowledge of trigonometry from the Greeks?
- 19.7. Archimedes wrote a work called the *Sand-reckoner* to prove that the universe (as the Greeks pictured it) could be filled with a *finite* number of grains of sand. The necessity of doing so shows that the Greeks had the same psychological difficulties that all people have in distinguishing clearly between “infinite” and “very large.” In the following passage from a Jain work, a related issue is addressed, namely what is the largest nameable number?

Consider a trough whose diameter is of the size of the earth. Fill it up with white mustard seeds counting them one after another. Similarly, fill up with mustard seeds other troughs of the sizes of the various lands and seas. Still it is difficult to reach the highest enumerable number.

Should the infinite be thought of as in some sense “approximated” by a very large finite quantity, or is it qualitatively different? Is it possible to create a meaningful arithmetic in which there is a largest integer?

- 19.8.** Are there any clues in the cultural context of Indian mathematics that help to explain why it was the only ancient civilization to develop a system of numeration that was based on both the number 10 and place value, so that only 10 symbols were needed to write it?

From the *Vedas* to Aryabhata I

A unique feature of arithmetic in ancient India, pointed out by Plofker (2009, pp. 14–15), is the existence of names for very large powers of 10, going beyond any conceivable practical social or commercial need. One early poem, the *Valmiki Ramayana*, from about 500 BCE, explains the numeration system in the course of recounting the size of an army. The description uses special words for 10^7 , 10^{12} , 10^{17} , and many other denominations, all the way up to 10^{55} . An important part of the place-value notation we now use is the zero symbol for an empty place, which may have been invented in India before 200 BCE. [Plofker (2009, p. 16) notes that while the concept of an empty place can be found in early documents, there is no clear “paper trail” to the first mathematical documents where it is known to occur, and (Plofker 2009, p. 48) it may not have been part of the early place-value decimal system, which was being used by the third century CE.]

20.1. PROBLEMS FROM THE *SULVA SUTRAS*

We now examine some mathematical problems posed in the *Vedas*. These problems were sometimes connected with the construction of altars. Our source for most of this material is the book of Srinivasiengar (1967).

20.1.1. Arithmetic

Some of the arithmetic content of the *Sulva Sutras* consists of rules for finding Pythagorean triples of integers, such as (3, 4, 5), (5, 12, 13), (8, 15, 17), and (12, 35, 37). It is not certain what practical use these arithmetic rules had. The best conjecture is that they were part of religious ritual. A Hindu home was required to have three fires burning at three different altars. The three altars were to be of different shapes, but all three were to have the same area. These conditions led to certain Diophantine-type problems, a particular case of which is the generation of Pythagorean triples, so as to make one square integer equal to the sum of two others.

One class of mathematical problems associated with altar building involves an altar of prescribed area having several layers. In one problem from the *Bodhayana Sutra* the altar is to have five layers of bricks, each layer containing 21 bricks. Now one cannot simply divide

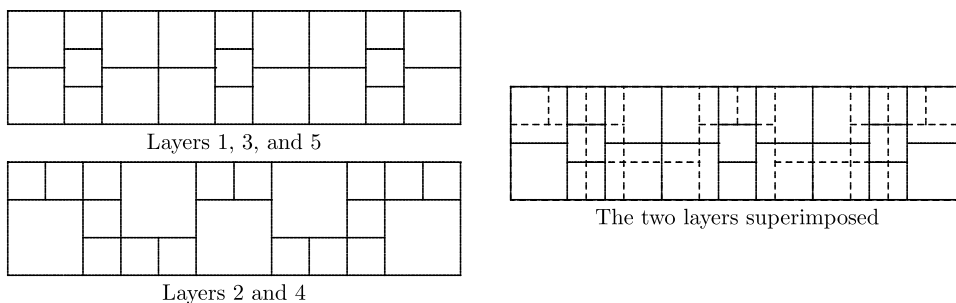


Figure 20.1. Construction of a brick altar with the same number of bricks and the same area in each of five layers.

a pile of 105 identical bricks into five layers and pile them up. Such a structure would not be stable. It is necessary to stagger the edges of the bricks. Thus, so that the outside of the altar will not be jagged, it is necessary to have at least two different sizes of bricks. The problem is to decide how many different sizes of bricks will be needed and how to arrange them. Assuming an area of one square unit—actually the unit is one square *vyayam*, a little over 6 square meters—the author suggests using three kinds of square bricks, of areas $\frac{1}{36}$, $\frac{1}{16}$, and $\frac{1}{9}$ square unit. The first, third, and fifth layers are to have 9 of the first kind and 12 of the second. The second and fourth layers get 16 of the first kind and 5 of the third. One way to arrange these layers so as to stagger the gaps in successive layers is shown in Fig. 20.1.

20.1.2. Geometry

The geometric content of the *Sulva Sutras* encompasses some of the transformation-of-area constructions such as we have seen in Euclid's *Elements*. The Pythagorean theorem is given, along with constructions for finding the side of a square equal to a rectangle, or the sum or difference of two other squares. The quadrature of a rectangle resembles the one found in Proposition 5 of Book 2 of Euclid rather than Euclid's construction of the mean proportional in Book 6, which is equivalent to it.

The Pythagorean theorem is not given a name, but is stated as the fact that “the diagonal of a rectangle produces both [areas] which its length and breadth produce separately.” It is interesting that the problem of doubling a square, which might have led to the discovery of this theorem, produces a figure in the shape of one of the altars discussed in the *Vedas*. Is it merely a coincidence that the problem of doubling the cube was said by the Greeks to have been inspired by an attempt to double the size of an altar?

The Hindu method of constructing of a square equal to a given rectangle (see Fig. 20.2) is as follows. Let $ABCD$ be the given rectangle, with AD longer than AB . Mark point E on AD so that $AE = AB$, and mark F on BC so that $BF = AB$. Draw EF , obtaining the square $ABFE$. Let G be the midpoint of ED and let H be the midpoint of FC . Draw GH and extend it to K so that $GK = AG$. Extend AB to L so that $AL = GK = AG$. Draw KL , obtaining the square $ALKG$. Extend EF to meet LK at M . Then the rectangle $ABCD$ equals the square $ALKG$ minus the square $HKMF$, since the rectangle $CDGH$ equals the rectangle $BLMF$. Next choose P on BH so that $PL = KL$. (This point can be located by drawing a circle with L as center and LK as radius, as shown in Fig. 20.2.) Draw the line from P perpendicular to LK meeting LK at Q . Then the square on LQ is the square on LP

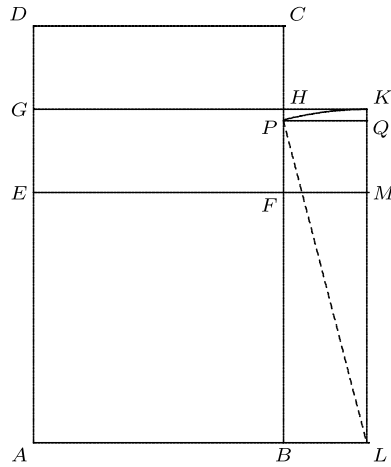


Figure 20.2. Quadrature of the rectangle in the *Sulva Sutras*.

minus the square on PQ . But since $PQ = HK$ and $LP = LK$, it follows that the square on LQ is precisely equal to the rectangle $ABCD$.

To construct a square equal to a multiple of a given square, say seven times as large as a square of side a , the *Katyayana Sutra* says to construct an isosceles triangle of base $6a$ and two sides equal to $4a$. The altitude, which is the perpendicular bisector of the base, will have length $a\sqrt{4^2 - 3^2} = \sqrt{7}a$, and hence will be the side of a square 7 times the original square.

The requirement of three altars of equal areas but different shapes would explain the interest in transformation of areas. Among other transformation of area problems, the authors of the *Vedas* considered the relative sizes of squares and circles. The *Bodhayana Sutra* states the problem of constructing a circle equal to a given square. The following approximate construction is given as the solution.

Let $ABCD$ be the square (see Fig. 20.3). From the center O of the square draw a circle with radius equal to OC . Let L be the midpoint of side BC , and let the radius through L meet the circle in the point E . Choose a point P on LE one-third of the way from L to E . The point P will lie on the circle with center at O equal to the square $ABCD$. In other

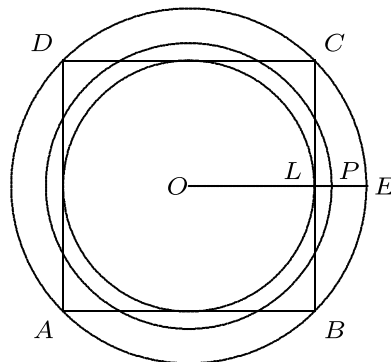


Figure 20.3. Rounding a square.

words, the radius of a circle equal to a given square is one-third the radius of the circle circumscribed about the square, plus two-thirds the radius of the circle inscribed in it. In contrast to the polygonal transformations just discussed, which were exact, this result is only approximate. In our terms, this construction gives a value for two-dimensional π of $18(3 - 2\sqrt{2})$, which is about 3.088.

20.1.3. Square Roots

The geometry of rectangles and right triangles leads naturally to the problem of handling numerical square roots, and accordingly the *Sulva Sutras* discuss a way of approximating them. The *Apastamba*, *Bodhayana*, and *Katyayana Sulva Sutras* (Plofker, 2009, p. 21) give the expression

$$1 + \frac{1}{3} + \frac{1}{3 \cdot 4} - \frac{1}{3 \cdot 4 \cdot 34}$$

for the diagonal of a square of side 1 (that is, $\sqrt{2}$). If this series represents successive approximations to $\sqrt{2}$, these approximations are $1, \frac{4}{3}, \frac{17}{12}, \frac{577}{408}$. The Mesopotamian approximation conjectured in Chapter 3 gives $1, 2, \frac{3}{2}, \frac{4}{3}, \frac{17}{12}, \frac{24}{17}, \dots$. One conjecture as to the origin of the present approximation is that it comes from the approximate equation

$$\sqrt{a^2 + r} \approx a + \frac{r}{2a} - \frac{(r/2a)^2}{2[a + r/2a]},$$

with $a = \frac{4}{3}$ and $r = \frac{2}{9}$. This approximation may be the source of similar rule given by the twelfth-century Moroccan mathematician Abu Bakr al-Hassar.

In the early seventh century, the mathematician Bhaskara I (ca. 600–ca. 680) expressed the opinion that the ratio of the circumference of a circle to its diameter cannot be exactly expressed. In Greek terms, the two are incommensurable; in our terms, π is irrational. By the late fourteenth century, the mathematician Madhava (ca. 1350–ca. 1425) gave a rule that expresses this ratio as an infinite series:

$$4\left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \dots\right).$$

This same series was given some 300 years later by Leibniz (Plofker, 2009, pp. 140, 224).

In the treatment of what we call irrational numbers, we see an instance in which the Greek insistence on logical correctness was a hindrance. The Greeks did not regard $\sqrt{2}$ as a number, since they could not express it exactly as a ratio and they knew that they could not. The Hindus may or may not have known of the impossibility of a rational expression for this number (they certainly knew that they did not *have* any rational expression for it); but, undeterred by the incompleteness of their knowledge, they proceeded to make what use they could of this number. This same “reckless” spirit served them well in the use of infinity and the invention of zero and negative numbers. They saw the usefulness of such numbers and either chose to live with or did not notice certain difficulties of a metaphysical character.

20.1.4. Jain Mathematics: The Infinite

Like Greek mathematics, Hindu mathematics has a prominent metaphysical component. This metaphysical aspect manifests itself in various ways—for example, in handling the infinite. Where the Greeks had regarded all reasoning as finite and accepted only a potential infinity, as shown by the method of exhaustion, the Hindus accepted an actual infinity and classified different kinds of infinities. This part of Hindu mathematics is particularly noticeable with the Jains. They classified numbers as enumerable, unenumerable, and infinite, and space as one-dimensional, two-dimensional, three-dimensional, and infinitely infinite. The first unenumerable number is the most unusual of these concepts. It is a finite number, but one can never describe it explicitly. The idea is to progress through the finite numbers 2, 3, 4, . . . in one's imagination until the “first unenumerable” number is reached. We can define it implicitly as the first positive integer that cannot be named. Does this mean the first number that no one *ever will* name (in the whole of human history), or the first number that *in principle could not* be named? The reader is invited to speculate.

20.1.5. Jain Mathematics: Combinatorics

The metaphysics of the Hindus, and especially the Jains, based on a classification of sentient beings according to the number of senses possessed, led them to a mathematical topic not discussed by the Greeks. The Hindus called it *vikalpa*, and we know it as *combinatorics*. The Sanskrit word *kalpa* has many meanings, among which are *possible*, *feasible*, and *ordered*. The prefix *vi-* corresponds roughly to the English prefix *dis-*, so that *vikalpa* may mean *distribution* in the sense of *arrangement*. The occurrence of the word in the present context probably derives from the *Kalpa Sutras*, a set of Jain verses.

Given that there are five senses and animals are to be classified according to the senses they possess, how many different classes will there be? A typical question might be, How many groups of three can be formed from a set of five elements? We know the answer, as did the early Jain mathematicians. In the *Bhagabati Sutra*, written about 300 BCE, the author asks how many philosophical systems can be formed by taking a certain number of doctrines from a given list of basic doctrines. After giving the answers for 2, 3, 4, etc., the author says that enumerable, unenumerable, and infinite numbers of things can be discussed, and, “as the number of combinations are formed, all of them must be worked out.”

The general process for computing combinatorial coefficients was known to the Hindus at an early date. Combinatorial questions seemed to arise everywhere for the Hindus, not only in the examples just given but also in a much earlier work on medicine that poses the problem of the number of different flavors that can be made by choosing subsets of six basic flavors (bitter, sour, salty, astringent, sweet, hot). The author gives the answer as $6 + 15 + 20 + 15 + 6 + 1$, that is, 63. We recognize here the combinatorial coefficients that give the subsets of various sizes that can be formed from six elements. The author did not count the possibility of no flavor at all.

Combinatorics also arose with the Hindus in the study of literature in the third century BCE, when Pingala gave a rule for finding the number of different words that could be formed from a given number of letters. This rule was written very obscurely, but a commentator named Halayudha in the tenth century CE explained it as follows. First draw a square. Below it and starting from the middle of the lower side, draw two squares. Then draw three squares below these, and so on. Write the number 1 in the middle of the top square and inside the first and last squares of each row. Inside every other square the number to be written is the

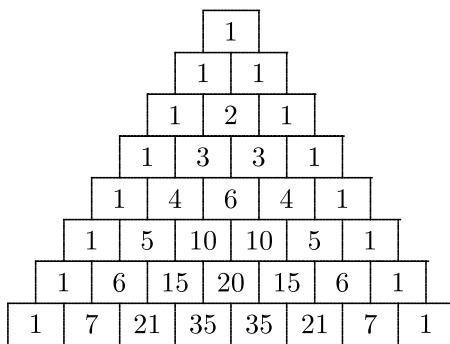


Figure 20.4. The Meru Prastara.

sum of the numbers in the two squares above it and overlapping it. This description of what we know as Pascal’s triangle was thus given in India 300 years before it was published in China and 700 years before Pascal. Moreover, it purports to be only a clarification of a rule invented 1200 years earlier! Its Sanskrit name is *Meru Prastara* (Fig. 20.4), which means the *Mount Meru staircase*.¹ The inspiration for the study of this figure was quite different in China and India. In China, it came about in connection with the extraction of roots and the solution of equations, whereas in India the inspiration was directly from the area of combinatorics.

According to Srinivasiengar (1967, p. 25), by the year 300 BCE Jain mathematicians understood certain cases of the laws of exponents. They could make sense of an expression like $a^{m/2^n}$, interpreting it as extracting the square root n times and then raising the result to the power m . The notation used was of course not ours. The power $\frac{3}{4}$, for example, was described as “the cube of the second square root.” That the laws of exponents were understood for these special values is attested by such statements as “the second square root multiplied by the third square root, or the cube of the third square root,” indicating an understanding of the equality $\sqrt{\sqrt{a}} \times \sqrt{\sqrt{\sqrt{a}}} = \left(\sqrt{\sqrt{\sqrt{a}}}\right)^3$, which we would write in exponential notation as

$$a^{1/4} a^{1/8} = a^{3/8} .$$

20.1.6. The Bakshali Manuscript

Some symbolic algebra can be found in the Bakshali manuscript. The symbol \ominus is used to denote an unknown quantity. One of the problems in the manuscript is written as follows, using modern number symbols and a transliteration of the Sanskrit into the Latin alphabet:

$$\begin{array}{ccccccc} \ominus & 5 & & \ominus & \ominus & 7 + & m\bar{u} & \ominus \\ 1 & 1 & yu & m\bar{u} & 1 & sa & 1 & 1 & 1 \end{array} .$$

This symbolism can be translated as follows: “A certain thing is increased by 5 and the square root is taken, giving [another] thing; and the thing is decreased by 7 and the square

¹In Hindu mythology, Mount Meru plays a role similar to that of Mount Olympus in Greek mythology. One Sanskrit dictionary gives this mathematical phrase as a separate entry.

root is taken, giving [yet another] thing.” In other words, we are looking for a number x such that $x + 5$ and $x - 7$ are both perfect squares. This problem is remarkably like certain problems in Diophantus. For example, Problem 11 of Book 2 of Diophantus’ *Arithmetica* is to add the same number to two given numbers so as to make each of them a square. If the two given numbers are 5 and -7 , this is *exactly* the problem stated here; Diophantus, however, did not use negative numbers.

The Bakshali manuscript also contains problems in linear equations, of the sort that has had a long history in elementary mathematics texts. For example, three persons possess seven thoroughbred horses, nine draft horses, and 10 camels, respectively. Each gives one animal to each of the others. The three are then equally wealthy. Find the (relative) prices of the three animals. Before leaping blindly into the set of two linear equations in three unknowns that this problem prescribes, we should take time to note that the problem can be solved by imagining the experiment actually performed. Suppose that these donations have been made and the three people are now equally wealthy. They will remain equally wealthy if each gives away one thoroughbred horse, one draft horse, and one camel. It follows that four thoroughbred horses, six draft horses, and seven camels are all of equal value. The problem has thereby been solved, and no actual algebra has been performed. Srinivasiengar (1967, p. 39) gives the solution using symbols for the unknown values of the animals, but does not assert that the solution is given this way in the manuscript itself.

20.2. ARYABHATA I: GEOMETRY AND TRIGONOMETRY

Chapter 2 of Aryabhata’s *Aryabhatiya* (Clark, 1930, pp. 21–50) is called *Ganitapada* (*Mathematics*). In Stanza 6 of this chapter, Aryabhata gives the correct rule for area of a triangle, but declares that the volume of a tetrahedron is half the product of the altitude and the area of the base. He says in Stanza 7 that the area of a circle is half the diameter times half the circumference, which is correct, and shows that he knew that one- and two-dimensional π were the same number. But he goes on to say that the volume of a sphere is the area of a great circle times its own square root. This would be correct only if three-dimensional π equaled $\frac{16}{9}$, very far from the truth! Plofker (2009, p. 126) discusses a suggested reinterpretation of this rule as applying to the surface area of the sphere rather than its volume and concludes that it will not do. In a way, this inaccurate result is surprising, since Aryabhata knew a very good approximation to one-dimensional π . In Stanza 10 he writes:

Add 4 to 100, multiply by 8, and add 62,000. The result is approximately the circumference of a circle of which the diameter is 20,000.

This procedure gives a value of one-dimensional π equal to 3.1416, which exceeds the true value by less than 0.01%.

Aryabhata also knew a method of surveying by sighting along the tops of two poles of equal height called *gnomons*. This same method was practiced in China. Whether this common method is a case of transmission or independent discovery is not clear. The rule given is illustrated by Fig. 20.5.

The distance between the ends of the two shadows multiplied by the length of the shadow and divided by the difference in length of the two shadows give the *koti*. The *koti* multiplied by the

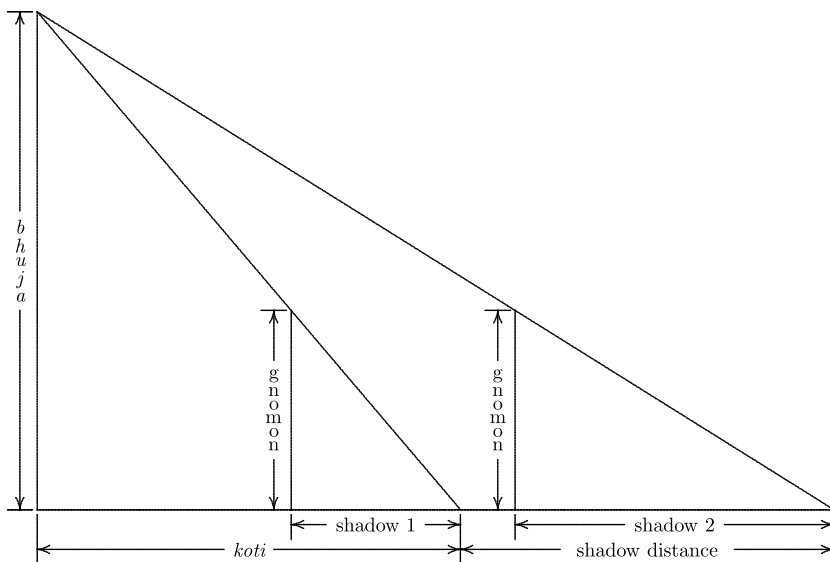


Figure 20.5. Aryabhata's method of surveying.

length of the gnomon and divided by the length of the shadow gives the length of the *bhuja*.
[Clark, 1930, p. 32]

20.2.1. Trigonometry

The inclusion of this method of surveying in the *Aryabhatiya* presents us with a small puzzle. As a method of surveying, it is not efficient. It would seem to make more sense to measure angles rather than using only right angles and measuring more lines. But angles are really not involved here. It is possible to have a clear picture of two mutually perpendicular lines without thinking “right angle.” The notion of angles in general as a species of mathematical objects—the figures formed by intersecting lines, which can be measured, added, and subtracted—appears to be a Greek innovation in the sixth and fifth centuries BCE, and it seems to occur only in plane geometry, not spherical, where arcs are used instead. Its origins may be in stonemasonry and carpentry, where regular polygons have to be fitted together. Astronomy probably also made some contribution. Since Aryabhata I was one of the pioneers of this trigonometry and was primarily an astronomer, it seems slightly inconsistent that he recommended this method of surveying. Perhaps the explanation is that measuring the sky and measuring the earth belong to different categories.

The earliest form of trigonometry was a table of correspondences between arcs and their chords. We know exactly how such a table was originally constructed, since we have already looked at Ptolemy's treatise on astronomy, written around 150 CE. Although this table fulfilled its purpose in astronomy, the chord is a cumbersome tool to use in studying plane geometry. For example, it was well known that in any triangle, the angle opposite the larger of two sides will be larger than the angle opposite the smaller side. But what is the exact, quantitative relation between the two sides and the two angles? The ratio of the sides has no simple relationship to the ratio of the angles or to the chords those angles subtend as

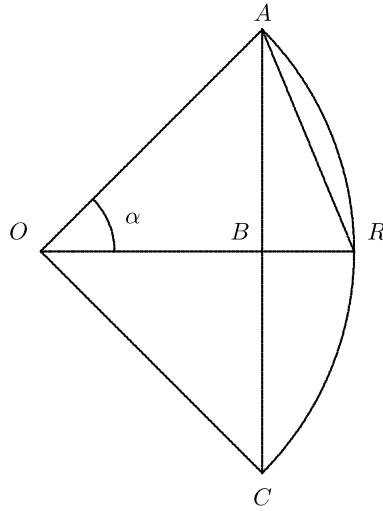


Figure 20.6. The “bowstring” diagram. The sine of the arc \widehat{AR} is the line AB . The tabular value of the sine is the number of minutes in an arc of the same length as AB .

central angles in a circle. But the sides are directly proportional to the chords of *twice* those central angles. In fact, as we have seen, Ptolemy was constantly working with the chord of twice an angle when he applied Menelaus’ theorem to solve spherical triangles. But he never used half of the chord of a doubled angle.

It was the Hindu astronomers who discovered that trigonometry is simpler if you use half of the chord of a doubled angle. Those half-chords are now called *sines*. In Fig. 20.6 the arc \widehat{AR} can be measured by either line AB or AR . Ptolemy chose AR and was led to the complications already mentioned. The Hindus preferred AB . We shall see that the Chinese word (*xian*) for the hypotenuse of a right triangle means *bowstring*. The Hindus used the Sanskrit term for a bowstring (*jya* or *jiva*) to mean the sine. The reason for the colorful language is obvious from Fig. 20.6.

To all appearances, then, trigonometry began to assume its modern form among the Hindus some 1500 years ago. A few reservations are needed, however. First, for the Hindu mathematicians the *sine* was not, as it is to us, a *ratio*. It was a *length*, and that physical dimension had to be taken into account in all computations. Second, the only Hindu concept corresponding approximately to our trigonometric functions were those of sine and cosine. The tangent, secant, cotangent, and cosecant were not included until much later. Third, the use of trigonometry was restricted to astronomy. As already pointed out, surveying, which is the other natural place to use trigonometry, did not depend on angle measurement.

Aryabhata used the sine function developed in the *Surya Siddhanta*, giving a table for computing its values at intervals of $225'$ ($3^\circ 45'$) of arc from 0° to 90° degrees and expressing these values in units of $1'$ of arc, rounded to the nearest integer, so that the sine of 90° , which is the radius of the circle, is 3438. In our terms, that number is, to the nearest integer, the number of minutes in one radian of arc, since the length of a one-radian arc is equal to the radius. The value is rounded up from 3437.75, however, and the use of that value in computations will yield better agreement with Aryabhata’s table. His sine of a given angle is 3438 times the number that we would call its sine.

Thus, the number 3438 is an artifact of the units chosen for the arcs and their sines. Since the unit of length for a sine is 1', the fact that the arc is closely approximated by the chord for small angles means that $\sin(\theta) \approx \theta$ for small arcs θ . This approximation holds within the limits of precision of the table up to 6° of arc when the arcs are also expressed in minutes, as Aryabhata does.

The $3\frac{3}{4}^\circ$ interval between entries suggests that the tables were computed independently of Ptolemy's work. If the Hindu astronomers had read Ptolemy, their tables of sines could easily have been constructed from his table of chords, and with more precision than is actually found. Almost certainly, this interval was reached by starting with an angle of 30°, whose sine was known to be half of the radius, then applying the formula for the sine of half an angle to get successively the sines of 15°, 7° 30', and finally 3° 45', which is 225'. Arybhata's table is actually a list of the *differences* of 24 successive sines at intervals of 225 minutes. Since one minute of arc is a very small quantity relative to the radius, these 24 values of the sine provide sufficient precision for the observational technology available at the time. Notice, however, that to calculate the sine of half of an angle θ one would have to carry out a computation equivalent to evaluating the cumbersome formula

$$\sin \frac{\theta}{2} = \sqrt{\frac{3437.75 - \sqrt{3437.75^2 - \sin^2 \theta}}{2}}.$$

It is therefore understandable that Aryabhata did not refine his table further. Aryabhata's list of sine differences is the following:

225, 224, 222, 219, 215, 210, 205, 199, 191, 183, 174,
164, 154, 143, 131, 119, 106, 93, 79, 65, 51, 37, 22, 7.

A comparison with a computer-generated table for the same differences reveals that Aryabhata's table is accurate, except that the sixth entry should be 211 instead of 210 and the eighth should be 198 instead of 199. But an error of only half of one percent is not critical, given the limited precision of Aryabhata's observations. It has been believed that this table of sine differences was computed by a recursive procedure, which can be described in our terms as follows (Clark, 1930, p. 29). Starting with $d_1 = 225$,

$$d_{n+1} = d_n - \frac{d_1 + \dots + d_n}{d_1},$$

where each term is rounded to the nearest integer after being calculated from this formula. Plofker (2009, p. 128), however, says that this interpretation of the text of the *Aryabhatiya* did not appear in any commentary until the fifteenth century; it is therefore not certain that this procedure is exactly what he meant. Moreover, a computer following this recursive instruction will generate a table that diverges from the one shown (see Problem 20.2).

Figure 20.7 shows a table of sine values that can be constructed on the basis of this table of differences. The first two columns give the arcs and their sines as implied by Aryabhata's table of differences. The third column converts Aryabhata's minutes to degrees and minutes. The fourth column gives the ratio of Aryabhata's sine to the radius (3438), which is what is nowadays called the sine. The last column gives the modern value of this sine and shows

Arc	Sine	Arc	Sine/radius	Sine [modern value]
225'	225'	3° 45'	0.065445	0.065403
450'	449'	7° 30'	0.130599	0.130526
675'	671'	11° 15'	0.195172	0.195090
900'	890'	15°	0.258871	0.258819
1125'	1105'	18° 45'	0.321408	0.321439
1350'	1315'	22° 30'	0.382490	0.382683
1575'	1520'	26° 15'	0.442118	0.442289
1800'	1719'	30°	0.500000	0.500000
2025'	1910'	33° 45'	0.555556	0.555570
2250'	2093'	37° 30'	0.608784	0.608761
2475'	2267'	41° 15'	0.659395	0.659346
2700'	2431'	45°	0.707097	0.707107
2925'	2585'	48° 45'	0.751891	0.751840
3150'	2728'	52° 30'	0.793485	0.793353
3375'	2859'	56° 15'	0.831588	0.831470
3600'	2978'	60°	0.866201	0.866025
3825'	3084'	63° 45'	0.897033	0.896873
4050'	3177'	67° 30'	0.924084	0.923880
4275'	3256'	71° 15'	0.947602	0.946930
4500'	3321'	75°	0.965969	0.965926
4725'	3372'	78° 45'	0.980803	0.980785
4950'	3409'	82° 30'	0.991565	0.991445
5175'	3431'	86° 15'	0.997964	0.997859
5400'	3438'	90°	1.000000	1.000000

Figure 20.7. Aryabhata’s table of sines (first two columns) and their modern equivalents (last three columns).

that it agrees up to three decimal places with the value in the fourth column. For use in the next chapter, we note that the sum of all of Aryabhata’s sines is 54, 233’.

Aryabhata applied the sine function to determine the altitude of the sun at a given hour of the day. The procedure is illustrated in Fig. 20.8 for an observer located at *O* in the northern hemisphere on a day in spring or summer. This figure shows a portion of the celestial sphere. The arc *RETSW* is the portion of the great circle in which the observer’s horizontal plane

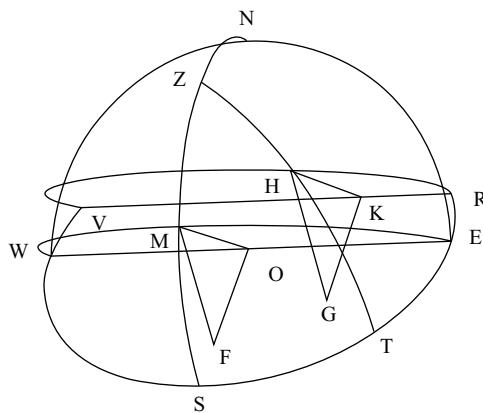


Figure 20.8. Finding the sun’s elevation at a given hour.

intersects the sphere. The sun will rise for this observer at the point R and set at the point V . The arc is slightly larger than a semicircle, since we are assuming a day in spring or summer. The chord RV runs from east to west. The sun will move along the small circle RHV at a uniform rate, and the plane of this circle is parallel to the equatorial circle EMW . (At the equinox, the day-circle RV coincides with the equatorial circle EW .) Aryabhata gave the correct formula for finding the radius of this day-circle in terms of the elevation of the sun above the celestial equator and the radius of the celestial sphere. That radius is the sine of the co-declination of the sun. Although Aryabhata had the concept of co-latitude, which served him in places where we would use the cosine function, for some reason he did not use the analogous concept of co-declination. As a result, he had to subtract the square of the sine of the declination from the square of the radius of the celestial sphere and then take the square root.

The point Z is the observer's zenith, M is the point on the celestial equator that is due south to the observer, and S is the point due south on the horizon, so that the arc \widehat{ZM} is the observer's terrestrial latitude, and the two arcs \widehat{ZN} and \widehat{MS} are both equal to the observer's co-latitude. The point H is the location of the sun at a given time, MF and HG are the projections of M and H respectively on the horizontal plane, and K is the projection of H on the chord RV . Finally, the great-circle arc HT , which runs through Z , is the altitude of the sun. The problem is to determine its sine HG in terms of lengths that can be measured.

Because their sides are parallel lines, the triangles MOF and HKG are similar, so that $MO : HK = MF : HG$. Hence we get

$$HG = \frac{HK \cdot MF}{MO}.$$

In this relation, MF is the sine of MS , that is, the sine of the observer's co-latitude, and MO is the radius of the celestial sphere. The line HK is, in a loose sense, the sine of the arc \widehat{RH} , which is proportional to the time elapsed since sunrise. It is perpendicular to the chord RV and would be a genuine sine if RV were the diameter of its circle. As it is, that relation holds only at the equinoxes. It is not certain whether Aryabhata meant his formula to apply only on the equinox, or whether he intended to use the word *sine* in this slightly inaccurate sense. Because the radius of the sun's small circle is never less than 90% of the radius of the celestial sphere, probably no observable inaccuracy results from taking HK to be a sine. In any case, that is the way Aryabhata phrased the matter:

The sine of the sun at any given point from the horizon on its day-circle multiplied by the sine of the co-latitude and divided by the radius is the [sine of the altitude of the sun] when any given part of the day has elapsed or remains. [Clark, 1930, p. 72]

20.2.2. The *Kuttaka*

Verses 32 and 33 of the *Aryabhatiya* contain a method known as the *kuttaka* (*pulverizer*) for solving problems related the "Chinese remainder theorem," which will be discussed in Chapter 22. Since the process was described more clearly by Brahmagupta, we reserve our discussion of it for the next chapter.

PROBLEMS AND QUESTIONS

Mathematical Problems

- 20.1.** Show that Aryabhata's list of sine differences can be interpreted in our language as the table whose n th entry is

$$3437.75 \left[\sin\left(\frac{n\pi}{48}\right) - \sin\left(\frac{(n-1)\pi}{48}\right) \right].$$

Here the angles are written in radian measure. Use a computer to generate this table for $n = 1, \dots, 24$, and compare the result with Aryabhata's table.

- 20.2.** If the recursive procedure said to have been used by Aryabhata is followed faithfully (as a computer can do), the result is the following sequence.

225, 224, 222, 219, 215, 210, 204, 198, 190, 181, 172,
162, 151, 140, 128, 115, 102, 88, 74, 60, 46, 31, 16, 1.

Compare this list with Aryabhata's list, and note the systematic divergence. These differences should be approximately 225 times the cosine of the appropriate angle. That is, $d_n \approx 225 \cdot \cos(225(n + 0.5) \text{ minutes})$. What does that fact suggest about the source of the systematic errors in the recursive procedure described by Aryabhata?

- 20.3.** Use Aryabhata's rule to compute the altitude of the sun above the horizon in London (latitude $51^\circ 32'$) at 10:00 AM (local solar time) on the vernal equinox. Assume that the sun rises at 6:00 AM on that day and sets at 6:00 PM.

Historical Questions

- 20.4.** Describe three kinds of geometric problems considered in the *Sulva Sutras*.
- 20.5.** How does the trigonometry used by Aryabhata I differ from what had been developed by Ptolemy four centuries earlier?
- 20.6.** Which geometric formulas given by Aryabhata I are accurate from the point of view of Euclidean geometry, and which are inaccurate?

Questions for Reflection

- 20.7.** Consider the problem posed by the Jain concept of the first unenumerable number. If this number is defined as the first number that no one ever will name, then in some sense it certainly exists (why?). But it will never be explicitly known to anyone, since, by definition, explicitly knowing a number means being able to name it. If it is defined as the first number that cannot even theoretically be named, another problem arises. Are there finite integers that cannot even theoretically be named? If there are, in what sense do they "exist"?

- 20.8.** Compare the conjecture given in the text as to the origin of the approximation for $\sqrt{2}$ with the following, due to a later commentator of 1500 CE. Assume that each side of the square is 12 units long. Then the diagonal has length $12\sqrt{2} = \sqrt{288} = \sqrt{17^2 - 1} \approx 17 - \frac{1}{34}$ [since $\sqrt{1-x} \approx 1 - (x/2)$]. It follows that $\sqrt{2} \approx \frac{17}{3 \cdot 4} - \frac{1}{3 \cdot 4 \cdot 34} = 1 + \frac{1}{3} + \frac{1}{3 \cdot 4} - \frac{1}{3 \cdot 4 \cdot 34}$. Which explanation seems more probable to you? Does either imply the other?
- 20.9.** Besides the sine function, we also use the tangent and secant and their cofunctions. What is the origin of the words *tangent* and *secant* (in Latin), and why are they applied to the objects of trigonometry?

Brahmagupta, the *Kuttaka*, and Bhaskara II

The present chapter is devoted to two mathematicians who lived 500 years apart. Of the two, the earlier (Brahmagupta) appears to be by far the more profound and original. Yet the second (Bhaskara II) is also well worth reading—as indeed are many later Hindu mathematicians, like the fifteenth-century Jyesthadeva, whose work we do not have space to discuss.

21.1. BRAHMAGUPTA’S PLANE AND SOLID GEOMETRY

Brahmagupta devotes five sections of Chapter 12 of the *Brahmasphutasiddhanta* to geometric results (Colebrooke, 1817, pp. 295–318). Like Aryabhata, he has a practical bent. In giving the common area formulas for triangles and quadrilaterals, he first gives a way of getting a “rough” value for the area: Take the product of the averages of the two pairs of opposite sides. (For this purpose a triangle counts as a quadrilateral having one side equal to zero.) In the days when calculation had to be done by hand, this was a quick approximation that worked well for quadrilaterals and triangles that are nearly rectangular (that is, tall, thin isosceles triangles). He also gave a formula that he says is exact, and this formula is a theorem commonly known as *Brahmagupta’s theorem*: *Half the sum of the sides set down four times and severally lessened by the sides, being multiplied together, the square root of the product is the area.* In our terms this rule says that the area of a quadrilateral of sides a , b , c , and d is $\sqrt{(s-a)(s-b)(s-c)(s-d)}$, where s is half of the sum of the lengths of the sides. The case when $d = 0$, which is a triangle, is what we call *Heron’s formula* and was discussed in Chapter 16. Brahmagupta did not mention the restriction that the quadrilateral must be a cyclic quadrilateral, that is, it must be inscribed in a circle.

Like Aryabhata, Brahmagupta knew that what we are calling one- and two-dimensional π were the same number. In Stanza 40, he says that when the diameter and the square of the radius respectively are multiplied by 3, the results are the “practical” circumference and area. In other words, $\pi = 3$ is a “practical” value. He also gives the “neat” (“exact”) value as $\sqrt{10}$. Since $\sqrt{10} = 3.1623$, this value is not an improvement on Aryabhata’s 3.1416 in terms of accuracy. If one had to work with π^2 , it might be more convenient. But π^2 occurs in very few contexts in mathematics, and none at all in elementary mathematics.

Section 5 of Chapter 12 of the *Brahmasphutasiddhanta* gives a rule for finding the volume of a frustum of a rectangular pyramid. In keeping with his approach of giving approximate rules, Brahmagupta says to take the product of the averages of the sides of the top and bottom in the two directions, then multiply by the depth. He calls this result the “practical measure” of the volume, and he knew that this simple rule gave a volume that was too small. To see why, imagine a frustum of height h with an $a \times b$ rectangle at the top and a proportional rectangle $ta \times tb$ at the bottom. The rule just stated would make this volume equal to $\frac{abh}{4}(1+t)^2$. Since the true volume is $\frac{abh}{3}(t^2+t+1)$, the difference is $\frac{t^2-2t+1}{4(t^2+t+1)}$ times the true volume. So, just as with his rule for triangles, if the pyramid has a very steep slope, so that t is close to 1, this value is reasonably accurate.

For his second approximation, which he called the “rough” volume, he took the average of the areas of the top and bottom and multiplied by the depth.¹ He also knew that this procedure gave a volume that was too large. In terms of the hypothetical frustum just introduced, it gives a volume of $\frac{abh}{2}(t^2+1)$, which is larger than the actual volume by $\frac{t^2-2t+1}{2(t^2+t+1)}$ of that volume. The actual volume lies between the “practical” volume and the “rough” volume, but where? From the explanation just given, it follows that the actual volume is obtained as a mixture of two parts “practical” and one part “rough.” Brahmagupta’s corrective procedure to give the “neat” (exact) volume was: Subtract the practical from the rough, divide the difference by three, and then add the quotient to the practical value. Although this rule seems rather roundabout, it is equivalent to the correct formula. It has some resemblance to the procedure given in the *Sulva Sutras* for constructing a circle equal to a square, which was discussed in Chapter 20.

21.2. BRAHMAGUPTA’S NUMBER THEORY AND ALGEBRA

Brahmagupta’s algebra is done entirely in words. For example (p. 279 of the Colebrooke translation), his recipe for the cube of a binomial is as follows:

The cube of the last term is to be set down; and, at the first remove from it, thrice the square of the last multiplied by the preceding; then thrice the square of the preceding term taken into that last one; and finally the cube of the preceding term. The sum is the cube.

In our terms, $(a+b)^3 = b^3 + 3b^2a + 3ba^2 + a^3$. This rule is used for finding successive approximations to the cube root, just as it was in China, as we shall see in the next two chapters. Similarly, in Section 4 (p. 346 of the Colebrooke translation), he tells how to solve a quadratic equation of the form $ax^2 + bx = c$:

Take the absolute number [the constant term c] from the side opposite to that from which the square and simple unknown are to be subtracted. To the absolute number multiplied by four times the [coefficient of the] square, add the square of the [coefficient of the] middle term; the square root of the same, less the [coefficient of the] middle term, being divided by twice the [coefficient of the] square is the [value of the] middle term.

¹This is the same procedure followed in the cuneiform tablet BM 85194, discussed above in Section 5.3 of Chapter 5.

Here the “middle term” is the unknown, and this statement is a very involved description of what we write as the quadratic formula:

$$x = \frac{\sqrt{4ac + b^2} - b}{2a} \quad \text{when} \quad ax^2 + bx = c.$$

Except for extracting cube roots of numbers, Brahmagupta does not consider equations of degree higher than 2.

Brahmagupta gave rules for handling sums of arithmetic progressions. (Aryabhata I had also done this.) He made systematic use of zero and negative numbers, giving the correct rules for manipulating them in the eighteenth chapter of the *Brahmasphutasiddhanta*. Brahmagupta devotes considerable space to the *pulverizer* (*kuttaka*) method of solving linear Diophantine equations, which was mentioned in the preceding chapter. Since this method is worth taking the time to master, we shall discuss it below. Before presenting it, however, we shall first discuss some of his other work in number theory and algebra.

21.2.1. Pythagorean Triples

Brahmagupta gave a method of creating Pythagorean triples of integers. In Chapter 12 of the *Brahmasphutasiddhanta* (p. 306 of the Colebrooke translation) he gives the rule that *the sum of the squares of two unlike quantities are the sides of an isosceles triangle; twice the product of the same two quantities is the perpendicular; and twice the difference of their squares is the base*. This rule amounts to the formula $(a^2 + b^2)^2 = (2ab)^2 + (a^2 - b^2)^2$, but it is stated as if the right triangle has been doubled by gluing another copy to the side of length $2ab$, thereby producing an isosceles triangle with base $2(a^2 - b^2)$, altitude $2ab$, and legs each $a^2 + b^2$. The relation stated is a purely geometric relation, showing (in our terms) that the sides and altitude of an isosceles triangle of any shape can be generated by choosing the two lengths a and b suitably. (In our terms, the equations $2(a^2 - b^2) = u$ and $2ab = v$ can be solved for a and b given any positive numbers u and v .)

21.2.2. Pell's Equation

Brahmagupta also considered generalizations of the problem of Pythagorean triples to a more general equation called² *Pell's equation* and written $Dx^2 - y^2 + 1 = 0$. He gives a recipe for generating a new equation of this form and its solutions from a given solution. The recipe proceeds by starting with two rows of three entries, which we shall illustrate for the case $D = 8$, which has the solution $x = 1, y = 3$. We write

$$\begin{array}{ccc} 1 & 3 & 1 \\ 1 & 3 & 1 \end{array}$$

²Erroneously so-called, according to Dickson (1920, p. 341), who asserts that Fermat had studied the equation earlier than John Pell (1611–1685). However, the MacTutor website at the University of St Andrews gives evidence that Euler's attribution of this equation to Pell was accurate. Everybody agrees that the solutions of the equation were worked out by Joseph-Louis Lagrange (1736–1813), not Pell.

The first column contains x , called the *lesser solution*, the second contains y , called the *greater solution*, and the third column contains the additive term 1. From these two rows a new row is created whose first entry is the sum of the cross-multiplied first two columns, that is $1 \cdot 3 + 3 \cdot 1 = 6$. The second entry is the product of the second entries plus 8 times the product of the first entries, that is $3 \cdot 3 + 8 \cdot 1 \cdot 1 = 17$, and the third entry is the product of the third entries. Hence we get a new row $6 \ 17 \ 1$, and indeed $8 \cdot 6^2 + 1 = 289 = 17^2$. In our terms, this says that if $8x^2 + 1 = y^2$ and $8u^2 + 1 = v^2$, then $8(xv + yu)^2 + 1 = (8xu + yv)^2$. More generally, Brahmagupta's rule says that if $ax^2 + d = y^2$ and $au^2 + c = v^2$, then

$$a(xv + yu)^2 + cd = (axu + yv)^2.$$

It is easy to verify that this rule is correct using modern algebraic notation. In his book (Weil, 1984), the number theorist André Weil (1906–1998) referred to the relation just written and the more general relation $(x^2 + Ny^2)(z^2 + Nt^2) = (xz \pm Nyt)^2 + N(xt \mp yz)^2$ as “Brahmagupta's identity” (his quotation marks).

However this relation was discovered, the *motivation* for studying it can be plausibly ascribed to a desire to approximate irrational square roots with rational numbers. Brahmagupta's rule with $c = d = 1$ gives a way of generating larger and larger solutions of the *same* Diophantine equation $ax^2 + 1 = y^2$. If you have two solutions (x, y) and (u, v) of this equation, which need not be different, then you have two approximations y/x and v/u for \sqrt{a} whose squares are, respectively, $1/x^2$ and $1/u^2$ larger than a . The new solution generated will have a square that is only $1/(xv + yu)^2$ larger than a . This aspect of the problem of Pell's equation turns out to have a close connection with its complete solution in the eighteenth century.

21.3. THE KUTTAKA

Brahmagupta gave a clearer explanation than Aryabhata had done of a method of solving what we call linear Diophantine equations, that is, equations of the form $ax = by + c$, where a , b , and c are given integers, and x and y unknown integers to be found. He applied this technique to computations involving astronomy and the calendar. We shall illustrate the method with such a computation, not one taken from Brahmagupta's work, but entirely in the spirit of that work.

It is well known that 19 solar years are almost exactly equal to 235 lunar months. Given that the moon was full on January 30, 2010, what is the next year in which it will be full on February 5? If we choose one 235th of a solar year as a unit of time T , so that $T \approx 1.55$ days, or 37 hours, 18 minutes, then one year is $235T$ and according to the fundamental relation, one month is $19T$. Since our unit of time T is about a day and a half, the period from January 30 to February 5, which is six days, amounts to $4T$, approximately. Thus we would like to find an integer number of years y and an integer number of months x such that

$$19xT = 235yT + 4T.$$

That is, we want x lunar months to exceed y solar years by $4T$. Canceling T , we see that we need to solve $19x = 235y + 4$. There are infinitely many solutions if there are any at all, since if (x_0, y_0) is a solution, so is $(x_0 + 235k, y_0 + 19k)$ for any integer k whatever.

Conversely, any two solutions (x_0, y_0) and (x_1, y_1) will differ by $(235k, 19k)$ for some k . Thus the problem is to find one solution. One way to do this is by trial and error: Just look at multiples of 235 until you find one that leaves a remainder of 15 when divided by 19 (since $15 + 4 = 19$). Thus you begin with

$$\begin{aligned} 235 &= 19 \times 12 + 7, \\ 2 \times 235 &= 19 \times 24 + 14, \\ 3 \times 235 &= 19 \times 37 + 2. \end{aligned}$$

Continuing in this way, you eventually get to $13 \times 235 = 19 \times 160 + 15 = 19 \times 161 - 4$, so that $19 \times 161 = 13 \times 235 + 4$. Thus, we can take $x = 161$, $y = 13$. In particular, the year will be $2010 + 13 = 2023$. (This is correct!) This method of finding a year on which the Moon will be full on a particular date is remarkably accurate, considering that the time period T is actually about 37 hours, and hence not exactly a day and a half. When it goes wrong in a short-term prediction, the moon will be full a day later or earlier in the predicted year.

Thus, the solution of linear Diophantine equations is not difficult. The only disadvantage to the method used above is the tedious trial-and-error procedure of getting one solution. That is where the method called the *kuttaka* (*pulverizer*) comes in. This technique shortens the labor of finding the first solution by a considerable amount, especially when the coefficients a , b , and c are large. Here are the steps you follow:

1. First, be sure the equation is written $ax = by + c$, where a and b are positive, and $b > a$. In other words, the constant term c needs to be on the same side of the equation as the larger coefficient, and the two coefficients must have the same sign. If they don't, replace y by a new variable $z = -y$, and then they will have the same sign. You can then multiply the equation by -1 if necessary to get them both positive. The constant term c may be positive or negative. (This "normalizing" is not absolutely essential, but experience shows that one has to be very careful when executing the *kuttaka*. The experienced user can handle variants in the method, but the beginner had better follow rigid rules.)
2. Second, perform the Euclidean algorithm procedure to find the greatest common divisor d of a and b . If it is larger than 1, then the expression $ax - by$ can only be a multiple of that greatest common divisor, so if c is *not* a multiple of it, there are no solutions, and you are finished.
3. If d divides c , take all of the *quotients*—*except the last one, which yields a remainder of 0*—and write them in a column. To illustrate with the equation $19x = 235y + 4$, which we considered above, we have the column

12
2
1
2

4. Next, augment that column with two more numbers at the bottom. The first one is c/d if the number of quotients is even and $-c/d$ if it is odd. In our case, we have an even number of quotients, and so we adjoin 4 at the bottom. The second additional number, which forms the bottom row of the array, is always 0. Thus we get the following column:

$$\begin{array}{c} 12 \\ 2 \\ 1 \\ 2 \\ 4 \\ 0 \end{array}$$

5. Now operate on this column, at each stage modifying the bottom entry and the entry two rows above it, as follows: The entry two rows above the bottom gets replaced by its product with the number below it, plus the number below that. Thus in this example, the first thing to do is to replace the 2 in the third row (counting the bottom row as row 1) by $2 \times 4 + 0 = 8$. The second part of the procedure is to erase the bottom number. Repeating this procedure until there are only two rows left yields

$$\begin{array}{ccccccc} 12 & & & & & & \\ & 12 & & & & & \\ 2 & & 12 & & & & \\ & 2 & & 12 & & & \\ 1 & & 2 & & 12 & & 396 \\ 2 & \rightarrow & 1 & \rightarrow & 32 & \rightarrow & 32 \\ & & 8 & & 12 & & \\ 4 & & & 8 & & & \\ & 4 & & & & & \\ 0 & & & & & & \end{array}$$

We should now have a solution, and indeed we do: $x = 396$, $y = 32$. It is not the smallest solution, however. We get a smaller one by subtracting 235 from x and 19 from y , yielding $x = 161$, $y = 13$.

This procedure needs to be practiced on some simple equations, such as $3x = 23y + 1$ and $17x = 11y - 5$, before the details will fall into place. The number of errors that can creep into this procedure is rather large. If the answer you get doesn't check when you put the values of x and y back into the equation, look for the following possible mistakes:

1. a and b must both be positive and their greatest common divisor d must also divide c if the equation is to have any solutions.
2. When the equation is written $ax = by + c$, you must have $b > a$.
3. Do *not* include the last quotient from the Euclidean algorithm in the column.
4. Adjoin c/d to the column of *quotients* (ignore the remainders in this algorithm), if you have an even number of quotients. (If c/d is negative, leave it negative in this case.) If the number of quotients is odd, adjoin $-c/d$. (If c/d is negative, make it positive in this case.)

These are the commonest sources of errors when carrying out this procedure. But of course, you also have to do the divisions with remainder carefully, avoiding computational errors.

21.4. ALGEBRA IN THE WORKS OF BHASKARA II

The *Lilavati* of Bhaskara II contains a collection of problems in algebra, which are sometimes stated as though they were intended purely for amusement. For example,

One pair out of a flock of geese remained sporting in the water, and saw seven times the half of the square-root of the flock proceeding to the shore, tired of the diversion. Tell me, dear girl, the number of the flock.

Like countless other unrealistic algebra problems that have appeared in textbooks over the centuries, this story is a way of posing to the student a specific quadratic equation, namely $\frac{7}{2}\sqrt{x} + 2 = x$, whose solution is $x = 16$.

21.4.1. The *Vija Ganita* (Algebra)

As mentioned in Chapter 19, Bhaskara II advertised his *Algebra* as an object of intellectual contemplation. We may agree that it fits this description. The problems, however, are just as fanciful as in the *Lilavati*. For example, the rule for solving quadratic equations is applied in the *Vija Ganita* (p. 212 of the Colebrooke translation) to find the number of arrows x that Arjuna (hero of the *Mahabharata*) had in his quiver, given that he shot them all, using $\frac{1}{2}x$ to deflect the arrows of his antagonist, $4\sqrt{x}$ to kill his antagonist's horse, six to kill the antagonist himself, three to demolish his antagonist's weapons and shield, and one to decapitate him. In other words, $x = \frac{1}{2}x + 4\sqrt{x} + 10$.

21.4.2. Combinatorics

Bhaskara gives a thorough treatment of permutations and combinations, which already had a long history in India. He describes combinatorial formulas such as

$$\binom{7}{3} = \frac{7 \cdot 6 \cdot 5}{1 \cdot 2 \cdot 3} = 35$$

by saying

Let the figures from one upward, differing by one, put in the inverse order, be divided by the same in the direct order; and let the subsequent be multiplied by the preceding and the next following by the foregoing. The several results are the changes by ones, twos, threes, etc.

He illustrates this principle by asking how many possible combinations of stressed and unstressed syllables there are in a six-syllable verse. His solution is as follows:

The figures from 1 to 6 are set down, and the statement of them, in direct and inverse order is

$$\begin{array}{cccccc} 6 & 5 & 4 & 3 & 2 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 \end{array}$$

The results are: changes with one long syllable, 6; with two 15; with three, 20; with four, 15, with five, 6; with all long, 1.

Bhaskara assures the reader that the same method can be used to find the permutations of all varieties of meter. He then goes on to develop some variants of this problem, for example,

A number has 5 digits and the sum of the digits is 13. If zero is not a digit, find the total number of possible numbers.

To solve this problem, you have to consider the possibility of two distinct digits (for example, 91111, 52222, 13333, 55111, 22333), three distinct digits (for example 82111, 73111) and count all the possible rearrangements of the digits.

Bhaskara reports that the initial syllables of the names for colors “have been selected by venerable teachers for names of values of unknown quantities, for the purpose of reckoning therewith.” He proceeds to give the rules for manipulating expressions involving such quantities; for example, the rule that we would write as $(-x - 1) + (2x - 8) = x - 9$ is written

$$\begin{array}{r} ya \dot{1} \quad ru \dot{1}, \\ ya 2 \quad ru \dot{8}, \\ \text{Sum} \quad ya 1 \quad ru \dot{9}, \end{array}$$

where the dots indicate negative quantities. The syllable *ya* is the first syllable of the word for *black*, and *ru* is the first syllable of the word for *species*.

Bhaskara gives the rule that we express as the quadratic formula for solving a quadratic equation by radicals, then goes on to give a criterion for a quadratic equation to have two (positive) roots. He also says (pp. 207–208 of the Colebrooke translation) that “if the solution cannot be found in this way, as in the case of cubic or quartic equations, it must be found by the solver’s own ingenuity.” That ingenuity includes some work that would nowadays be regarded as highly inventive, not to say suspect; for example (p. 214), Bhaskara’s solution of the equation

$$\frac{(0(x + \frac{1}{2}x))^2 + 2(0(x + \frac{1}{2}x))}{0} = 15.$$

Bhaskara warns that multiplying by zero does not make the product zero, since further operations are to be performed. Then he simply cancels the zeros, saying that, since the multiplier and divisor are both zero, the expression is unaltered. The result is the equation we would write as $\frac{9}{4}x^2 + 3x = 15$. Bhaskara clears the denominator and writes the equivalent of $9x^2 + 12x = 60$. Even if the multiplication by zero is interpreted as multiplication by a nonzero expression that is *tending to zero*, as a modern mathematician would like to

do, this cancelation is not allowed, since the first term in the numerator is a higher-order infinitesimal than the second. Bhaskara is handling 0 here as if it were 1. Granting that operation, he does correctly deduce, by completing the square (adding 4 to each side), that $x = 2$.

Bhaskara says in the *Vija Ganita* that a nonzero number divided by zero gives an infinite quotient.

This fraction $\left[\frac{3}{0}\right]$, of which the denominator is cipher, is termed an infinite quantity. In this quantity consisting of that which has cipher for its divisor, there is no alteration, though many be inserted or extracted; as no change takes place in the infinite and immutable GOD [Vishnu], at the period of the destruction or creation of worlds, though numerous orders of beings are absorbed or put forth.

By the time of Bhaskara, the distinction between a rational and an irrational square root was well known. The Sanskrit word for an irrational root is *carani*, according to the commentator Krishna (Plofker, 2009, p. 145), who defines it as a number, “the root of which is required but cannot be found without residue.” Bhaskara gives rules such as $\sqrt{8} + \sqrt{2} = \sqrt{18}$ and $\sqrt{8} - \sqrt{2} = \sqrt{2}$.

21.5. GEOMETRY IN THE WORKS OF BHASKARA II

In his work *Siddhanta Siromani (Crest Jewel of the Siddhantas)*, written in 1150, Bhaskara tackled the extremely difficult problem of finding the area of a sphere. As we have seen (Section 7.2 of Chapter 7), the Egyptians had deduced correctly that the area of a hemisphere is twice the area of its circular base, and (Section 14.2 of Chapter 14) Archimedes had proved rigorously that the surface of a sphere is four times the equatorial disk it contains. In order to achieve that result, Archimedes had to make use of the method of exhaustion, which can be seen as an anticipation of integral calculus. Something similar can be said about Bhaskara’s approach, which was numerical and based on Aryabhata’s trigonometry, in contrast to the metric-free approach used by Archimedes. The discussion we are about to give is based on the exposition of this result given by Plofker (2009, pp. 196–201).

As was stated in the previous chapter, in constructing his table of sine differences, Aryabhata I chose $225'$ of arc as the constant difference, dividing the arc of one quadrant of a circle of radius $3438'$ into 24 equal pieces. Bhaskara II started from that point, dividing a complete great circle of a sphere—which we can think of as the equator—into 96 equal pieces. Each of these pieces is regarded as one unit of length. He then imagined the lines of longitude drawn through these 96 points running from pole to pole, thereby partitioning the sphere into 96 mutually congruent sectors. In each sector, he then imagined the circles of latitude drawn, dividing each quadrant of a line of longitude into 24 equal arcs between the pole and the equator, 48 between the two poles. Thus the surface of the sphere was partitioned into $96 \times 48 = 4608$ regions, 192 of which (those having a vertex at one of the poles) are curvilinear triangles, and the other 4416 of which are curvilinear trapezoids. A set of 20 of these trapezoids lying just above the equator is shown in Fig. 21.1. Since they are very small, one can imagine that they actually are planar triangles and trapezoids. For a typical trapezoid whose upper and lower edges are at co-latitudes $(k - 1) \times 225'$ and $k \times 225'$ —these are the distances along a line of longitude from the pole—the lengths of these edges are proportional to the radii of those circles of latitude. In terms of the unit

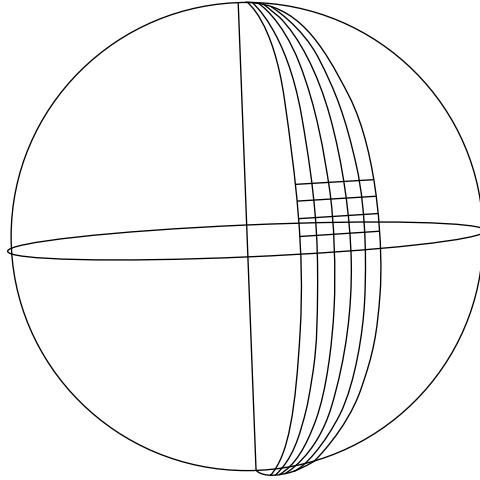


Figure 21.1. Bhaskara’s “polygonal” method of getting the area of a sphere.

of length (1′) chosen by Aryabhata, the radius r of the circle at co-latitude $k \times 225'$ is $\sin(k \times 225')$; that is, it is given in the second column of the table of sines displayed in the previous chapter. The length of the arc of that circle inside the sector is $225'$. However, these are minutes of arc *on the circle of latitude, not on the sphere*. A minute of arc on a circle of radius r is $\frac{r}{R}$ minutes of arc on a great-circle of a sphere of radius R . Thus, the portion of the circle of latitude of radius r inside each sector has length $\frac{225r}{R}$ minutes of spherical arc, where $R = 3438$. Since Bhaskara’s unit of length is 225 of Aryabhata’s units, we need to divide by 225. Altogether then, the length of that arc inside a given sector at co-latitude $k \times 225'$ is $\frac{\sin(k \times 225')}{R}$. The area of each trapezoid (again, treated as if it were a plane trapezoid) is numerically equal to the average of these lengths for the upper and lower edges, since each trapezoid has altitude equal to one unit. All we have to do then is sum up the areas of the 4416 trapezoids and the 192 triangles in order to find the area of the sphere. This is done most easily by finding the area of a half-sector and doubling it. To find the area of each triangle or trapezoid between the pole and the equator in a given sector, one has only to take the average of the lengths of the opposite sides (counting the “side” at the pole as having length 0), multiply by the altitude, and then add up the results. Thus, we need to find

$$\frac{1}{3438} \times \left(\frac{0 + \sin(225')}{2} + \frac{\sin(225') + \sin(2 \times 225')}{2} + \frac{\sin(2 \times 225') + \sin(3 \times 225')}{2} + \dots + \frac{\sin(23 \times 225') + \sin(24 \times 225')}{2} \right).$$

Bhaskara saw how this sum could be rewritten to eliminate the 2 in each denominator, except for the very last term. He evaluated it by adding up all of the sines in the table and then subtracting half of the last one. This was a simple exercise in arithmetic, and we noted

in the previous chapter that the sines in the table add up to 54, 233. Therefore, the area of a half-sector is numerically

$$\frac{(54233 - 1719)}{3438} = \frac{52514}{3438} = 15.27457.$$

(The number 1719 is half of the sine of a 90° arc.) A full sector is then twice this amount, or 30.54916. Bhaskara observed that this is, within the limits of precision, precisely the diameter of the sphere, since $\frac{96}{\pi} \approx 30.5577$. Thus, it seems that if a sphere is partitioned into sectors of unit opening, the area of each sector in square units is numerically equal to the length of the diameter. Since the total area of the sphere is 96 times the area of this sector—that is, it is this area times the number of units of length in the circumference, Bhaskara concluded (correctly) that *the area of a sphere [in square units] is equal to its diameter times its circumference*. Bhaskara would have had to construct a finer table of sines in order to test the result with a smaller unit of length (by partitioning the sphere into more than 4608 regions). As a practical matter, since the actual diameter is about 0.01 units larger than the value he used, while he had the circumference correct, he would have gotten a numerical value for the area that is too small by 1%. In our terms, his unit of area was $\frac{\pi^2 R^2}{4n^2}$, and the numerical approximation that he used for the area was

$$c_n \pi R^2,$$

where

$$c_n = \frac{\sqrt{2}\pi}{n} \left(\frac{\sin\left(\frac{(n-1)\pi}{4n}\right)}{\sin\left(\frac{\pi}{4n}\right)} + \frac{1}{2} \right).$$

The accurate value of c_n would be 4. Bhaskara's procedure amounts to taking $n = 24$. By direct computation, we get $c_{24} \approx 3.96023$, which is, as already noted, 1% too small.

That Bhaskara understood the principle of infinitesimal approximation is shown by another of his results, in which he says that the difference between two successive sines in the table, that is, $\sin((k+1) \times 225') - \sin(k \times 225')$, is $225 \cos(k \times 225')/R$ (where $R = 3438$). This result seems to prefigure the infinitesimal relation that calculus books write as

$$\Delta(\sin(x)) \approx d(\sin(x)) = \cos(x) dx.$$

PROBLEMS AND QUESTIONS

Mathematical Problems

- 21.1.** Given the Pell equation $y^2 - 11x^2 = 1$, which has solutions $x = 3, y = 10$ and $x = 60, y = 199$, construct a third solution and use it to get an approximation to $\sqrt{11}$.
- 21.2.** Solve Bhaskara's problem of finding the number of positive integers having five nonzero digits whose sum is 13.

- 21.3. Use the *kuttaka* to solve the equation $24x = 57y + 15$. Find the smallest positive integers x and y that satisfy this equation.

Historical Questions

- 21.4. How accurate are the rules given by Brahmagupta for computing areas and volumes?
- 21.5. What topics in number theory not discussed by Euclid and Nicomachus can be found in the works of Hindu mathematicians?
- 21.6. How did Bhaskara II treat division by zero?

Questions for Reflection

- 21.7. How practical was it to use the *kuttaka* to compute the dates of future conjunctions of the heavenly bodies (for example, eclipses)? Does this technique yield accurate and reliable results? What might go wrong in a given practical application?
- 21.8. What justification does Bhaskara II offer for the problems in the *Lilavati*? Does he live up to his advertising?
- 21.9. Compare the trigonometries developed by Ptolemy and the Hindu mathematicians with each other and with trigonometry as we know it today. What significant differences are there between any two of them?

Early Classics of Chinese Mathematics

The name *China* refers to a region unified under a central government but whose exact geographic extent has varied considerably over the 4000 years of its history. To frame our discussion, we shall sometimes refer to the following dynasties:

1. *The Shang Dynasty* (sixteenth to eleventh centuries BCE). The Shang rulers controlled the northern part of what is now China and had an extensive commercial empire.
2. *The Zhou Dynasty* (eleventh to eighth centuries BCE). The Shang Dynasty was conquered by people from the northwest known as the Zhou. The great Chinese philosophers known in the West as Confucius, Mencius, and Lao-Tzu lived and taught during the several centuries of disorder that came after the decay of this dynasty.
3. *The Period of Warring States* (403–221 BCE) and the *Qin Dynasty* (221–206 BCE). Warfare was nearly continuous in the fourth and third centuries BCE, but in the second half of the third century the northwestern border state of Qin gradually defeated all of its rivals and became the supreme power under the first Qin emperor. The name *China* is derived from the Qin.
4. *The Han Dynasty* (206 BCE–220 CE). The empire was conquered shortly after the death of the Qin emperor by people known as the Han, who expanded their control far to the south, into present-day Viet Nam, and established a colonial rule in the Korean peninsula. Contact with India during this dynasty brought Buddhism to China for the first time. According to Mikami (1913, pp. 57–58), mathematical and astronomical works from India were brought to China and studied. Certain topics, such as combinatorics, are common to both Indian and Chinese treatises, but “there is nothing positive that serves as an evidence of any actual Indian influence upon the Chinese mathematics.”
5. *The Tang Dynasty* (seventh and eighth centuries). The Tang Dynasty was a period of high scholarship, in which, for example, block printing was invented.
6. *The Song Dynasty* (960–1279). The period of disorder after the fall of the Tang Dynasty ended with the accession of the first Song emperor. Confucianism underwent a resurgence in this period, supplementing its moral teaching with metaphysical speculation. Scientific treatises on chemistry, zoology, and botany were written, and the Chinese made great advances in algebra.

7. *The Mongol conquest and the closing of China.* The Song Dynasty was ended in the thirteenth century by the Mongol conquest under the descendants of Genghis Khan, whose grandson Kublai Khan was the first emperor of the dynasty known in China as the Yuan. As the Mongols were Muslims, this conquest brought China into contact with the intellectual achievements of the Muslim world. Knowledge flowed both ways, and the sophisticated Chinese methods of root extraction seem to be reflected in the works of later Muslim scholars, such as the fifteenth-century mathematician al-Kashi. The vast Mongol Empire facilitated East–West contacts, and it was during this period that Marco Polo (1254–1324) made his famous voyage to the Orient.
8. *The Ming Dynasty* (fourteenth to seventeenth centuries). While the Mongol conquest of Russia lasted 240 years, the Mongols were driven out of China in less than a century by the first Ming emperor. During the Ming Dynasty, Chinese trade and scholarship recovered rapidly. The effect of the conquest, however, was to encourage Chinese isolationism, which became the official policy of the later Ming emperors during the period of European expansion. The first significant European contact came in the year 1582, when the Italian Jesuit priest Matteo Ricci (1552–1610) arrived in China. The Jesuits were particularly interested in bringing Western science to China to aid in converting the Chinese to Christianity. They persisted in these efforts despite the opposition of the emperor. The Ming Dynasty ended in the mid-seventeenth century with conquest by the Manchus.
9. *The Ching (Manchu) Dynasty* (1644–1911). After two centuries of relative prosperity the Ching Dynasty suffered from the depredations of foreign powers eager to control its trade. Perhaps the worst example was the Opium War of 1839–1842, fought by the British in order to gain control of the opium trade. From that time on, Manchu rule declined. In 1900, the Boxer Rebellion against the Western occupation was crushed and the Chinese were forced to pay heavy reparations. In 1911 the government disintegrated entirely, and a republic was declared.
10. *The twentieth century.* The establishment of a republic in China did not quell the social unrest, and there were serious uprisings for several decades. China suffered badly from World War II, which began with a Japanese invasion in the 1930s. Although China was declared one of the major powers when the United Nations was formed in 1946, the Communist revolution of 1949 drove its leader Chiang Kai-Shek to the island of Taiwan. China is now engaged in extensive cultural and commercial exchanges with countries all over the world and hosted the International Congress of Mathematicians in 2002. Its mathematicians have made outstanding contributions to the advancement of mathematics, and Chinese students are welcomed at universities in nearly every country.

22.1. WORKS AND AUTHORS

Mathematics became a recognized and respected area of intellectual endeavor in China more than 2000 years ago. That its origins are at least that old is established by the existence of books on mathematics, at least one of which was probably written before the order of the Emperor Shih Huang-Ti in 213 BCE that all books be burned.¹ A few books survived or

¹The Emperor was not hostile to learning, since he did not forbid the *writing* of books. Apparently, he just wanted to be remembered as the emperor in whose reign everything began.

were reconstituted after the brief reign of Shih Huang-Ti, among them the mathematical classic just alluded to. This work and three later ones now exist in English translation, with commentaries to provide the proper context for readers who are unfamiliar with the history and language of China. Under the Tang dynasty, a standardized educational system came into place for the training of civil servants, based on literary and scientific classics, and the works listed below became part of a mathematical curriculum known as the *Suan Jing Shishu* (*Ten Canonical Mathematical Classics*—there are actually 12 of them). Throughout this long period, mathematics was cultivated together with astronomy both as an art form and for practical application in the problem of obtaining an accurate lunisolar calendar. In addition, many problems of commercial arithmetic and civil administration appear in the classic works.

22.1.1. The *Zhou Bi Suan Jing*

The early treatise alluded to above, the *Zhou Bi Suan Jing*, has been known in English as the *Arithmetic Classic of the Gnomon and the Circular Paths of Heaven*. A recent study and English translation has been carried out by Christopher Cullen of the University of London (1996). According to Cullen, the title *Zhou Bi* could be rendered as *Gnomon of the Zhou*. The phrase *suan jing* occurs in the titles of several early mathematical works; it means *mathematical treatise* or *mathematical manual*. According to a tradition, the *Zhou Bi Suan Jing* was written during the Western Zhou dynasty, which overthrew the earlier Shang dynasty around 1025 BCE and lasted until 771 BCE. Experts now believe, however, that the present text was put together during the Western Han dynasty, during the first century BCE, and that the commentator Zhao Shuang, who wrote the version we now have, lived during the third century CE, after the fall of the Han dynasty. However, the astronomical information in the book could only have been obtained over many centuries of observation and therefore must be much earlier than the writing of the treatise.

As the traditional title shows, the work is concerned with astronomy and surveying. The study of astronomy was probably regarded as socially useful in two ways: (1) It helped to regulate the calendar, a matter of great importance when rituals were to be performed; (2) it provided a method of divination (astrology), also of importance both for the individual and for the state. Surveying is of use in any society where it is necessary to erect large structures such as dams and bridges and where land is often flooded, requiring people to abandon their land holdings and reclaim them later.

These applications make mathematics useful in practice. However, the preface, written by the commentator Zhao Shuang, gives a different version of the motive for compiling this knowledge. Apparently a student of traditional Chinese philosophy, he had realized that it was impossible to understand fully all the mysteries of the changing universe. He reports that he had looked into this treatise while convalescing from an illness and had been so impressed by the acuity of the knowledge it contained that he decided to popularize it by writing commentaries to help the reader over the hard parts, saying, “Perhaps in time, gentlemen with a taste for wide learning may turn their attention to this work” (Cullen, 1996, p. 171). Here we see mathematics being praised simply because it confers understanding where ignorance would otherwise be; it is regarded as one of the liberal arts, to be studied by a leisured class of *gentlemen* scholars, people fortunate enough to be free of the daily grind of physical labor that was the lot of the majority of people in all countries until very recent times.

22.1.2. The *Jiu Zhang Suan Shu*

Another ancient Chinese treatise, the *Jiu Zhang Suan Shu*, meaning *Nine Chapters on the Mathematical Art*,² has been partly translated into English, with commentary, by Lam (1994). A corrected and commented edition was published in Chinese in 1992, assembled by Guo (1992). This work has been called *the* classic Chinese mathematical treatise, since commentaries were written on it for centuries, and it had a large influence on the development of mathematics in Korea and Japan. It reflects the state of mathematics in China in the later Han dynasty, around the year 100 CE. The nine chapters that give this monograph its name contain 246 applied problems of a sort useful in teaching how to handle arithmetic and elementary algebra and how to apply them in commercial and administrative work. In that respect, it offers many parallels with the Rhind papyrus. The nine chapters have no prefaces in which the author explains their purpose, and so we must assume that the purpose was the obvious one of training people engaged in surveying, administration, and trade. Some of the problems are practical, explaining how to find areas, convert units of length and area, and deal with fractions and proportions. Yet when we analyze the algebraic parts of this work, we shall see that it contains impractical puzzle-type problems leading to systems of linear equations and resembling problems that have filled up algebra books for centuries. Such problems are apparently intended to train the mind in algebraic thinking.

22.1.3. The *Sun Zi Suan Jing*

Another early treatise, the *Sun Zi Suan Jing* or *Mathematical Classic of Sun Zi*, was written several centuries after the *Jiu Zhang Suan Shu*. This work begins with a preface praising the universality of mathematics for its role in governing the lives of all creatures and placing it in the context of Chinese philosophy and among the six fundamental arts (decorum, music, archery, charioteership, calligraphy, and mathematics).

The preface makes it clear that mathematics is appreciated both as a practical skill in life and as an intellectual endeavor. The practicality comes in the use of compasses and gnomons for surveying and in the use of arithmetic for computing weights and measures. The intellectual skill, however, is emphasized. Mathematics is valued because it trains the mind. “If one neglects its study, one will not be able to achieve excellence and thoroughness” (Lam and Ang, 1992, p. 151).

As in the quotation from the commentary on the *Zhou Bi Suan Jing*, we find that an aura of mystery and “elitism” surrounds mathematics. It is to be pursued by a dedicated group of initiates, who expect to be respected for learning its mysteries, as theologians were during the Middle Ages in the West. At the same time, mathematics has a practical value that is also respected.

22.1.4. Liu Hui. The *Hai Dao Suan Jing*

The fall of the Han Dynasty in the early third century gave rise to three separate kingdoms in the area now known as China. The north-central kingdom is known as the Kingdom of Wei. There, in the late third century CE, a mathematician named Liu Hui (ca. 220–280) wrote a commentary on the final chapter of the *Jiu Zhang Suan Shu*. This chapter is devoted to

²Martzloff (1994) translates this title as *Computational Prescriptions in Nine Chapters*.

the theorem we know as the Pythagorean theorem, and Liu Hui's book, the *Hai Dao Suan Jing* (*Sea Island Mathematical Classic*), shows how to use pairs of similar right triangles to measure inaccessible distances. The name of the work comes from the first problem in it, which is to find the height of a mountain on an offshore island and the distance to the base of the mountain. The work consists of nine problems in surveying that can be solved by the algebraic techniques practiced in China at the time. A translation of these problems, a history of the text itself, and commentary on the mathematical techniques can be found in the paper by Ang and Swetz (1986).

22.1.5. Zu Chongzhi and Zu Geng

According to Li and Du (1987, pp. 80–82), fifth-century China produced two outstanding mathematicians, father and son. Zu Chongzhi (429–501) and his son Zu Geng (ca. 450–520) were geometers who used a method resembling what is now called *Cavalieri's principle* for calculating volumes bounded by curved surfaces. The elder Zu was also a numerical analyst, who wrote a book on approximation entitled *Zhui Shu* (*Method of Interpolation*), which became for a while part of the classical curriculum. However, this book was apparently regarded as too difficult for nonspecialists, and it was dropped from the curriculum and lost. Zu Geng continued working in the same area as his father and had a son who also became a mathematician.

22.1.6. Yang Hui

We now leave a considerable (700-year) gap in the story of Chinese mathematics and come to Yang Hui (ca. 1238–1298), the author of a number of mathematical texts. According to Li and Du (1987, pp. 110, 115), one of these was *Xiangjie Jiuzhang Suan Fa* (*Detailed Analysis of the Mathematical Rules in the Jiu Zhang Suan Shu*), a work of 12 chapters, one on each of the nine chapters of the *Jiu Zhang Suan Shu*, plus three more containing other methods and more advanced analysis. In 1274 and 1275 he wrote two other works, which were later collected in a single work called the *Yang Hui Suan Fa* (*Yang Hui's Computational Methods*). In these works he discussed not only mathematics but also its pedagogy, advocating real understanding over rote learning.

22.1.7. Cheng Dawei

In the later Ming dynasty, a governmental administrator named Cheng Dawei (1533–1606) applied his mind to the solution of problems using the abacus. In 1592 he wrote a book entitled *Suan Fa Tong Zong* (*General Source of Computational Methods*), containing nearly 600 problems on a huge variety of topics, including magic squares and even more arcane subjects.

22.2. CHINA'S ENCOUNTER WITH WESTERN MATHEMATICS

Jesuit missionaries who entered China during the late sixteenth century brought with them some mathematical works, including Euclid's *Elements*, the first six books of which the missionary Matteo Ricci and the Chinese scholar Xu Guangchi (1562–1633) translated into Chinese (Li and Du, 1987, p. 193). The version of Euclid that they used, a Latin translation by

the German Jesuit Christopher Clavius (1538–1612) bearing the title *Euclidis elementorum libri XV* (*The Fifteen Books of Euclid's Elements*), is still extant, preserved in the Beijing Library. This book aroused interest in China because it was the basis of Western astronomy and therefore offered a new approach to the calendar and to the prediction of eclipses. According to Mikami (1913, p. 114), the Western methods made a correct prediction of a solar eclipse in 1629, which traditional Chinese methods got wrong. It was this accurate prediction that attracted the attention of Chinese mathematicians to Euclid's book, rather than the elaborate logical structure which is its most prominent distinguishing characteristic. Martzloff (1993) studied a commented (1700) edition of Euclid by the mathematician Du Zhigeng and noted that it was considerably abridged, omitting many proofs of propositions that are visually or topologically obvious. As Martzloff says, although Du Zhigeng retained the logical form of Euclid—that is, the definitions, axioms, postulates, and propositions—he neglected proofs, either omitting them entirely or giving only a fraction of a proof, “a fraction not necessarily containing the part of the Euclidean argument relevant to a given proposition and devoted to the mathematical proof in the proper sense of the term.” Du Zhigeng also attempted to synthesize the traditional Chinese classics, such as the *Jiu Zhang Suan Shu* and the *Suan Fa Tong Zong*, with works imported from Europe, such as Archimedes' treatise on the measurement of the circle. Thus in China, Western mathematics supplemented, but did not replace, the mathematics that already existed.

The first Manchu Emperor Kang Xi (1654–1722) was fascinated by science and insisted on being taught by two French Jesuits, Jean-François Gerbillon (1654–1707) and Joachim Bouvet (1656–1730), who were in China in the late 1680s. This was the time of the Sun King, Louis XIV, who was vying with Spain and Portugal for influence in the Orient. The two Jesuits were required to be at the palace from before dawn until long after sunset and to give lessons to the Emperor for four hours in the middle of each day (Li and Du, 1987, pp. 217–218).

Given the increasing contacts between East and West in the nineteenth century, some merging of ideas was inevitable. During the 1850s the mathematician Li Shanlan (1811–1882), described by Martzloff (1982) as “one of the last representatives of Chinese traditional mathematics,” translated a number of contemporary works into Chinese, including an 1851 calculus textbook of the American astronomer–mathematician Elias Loomis (1811–1889) and an algebra text by Augustus De Morgan (1806–1871). Li Shanlan had a power over formulas that reminds one in many ways of the twentieth-century Indian genius Srinivasa Ramanujan. One of his combinatorial formulas, stated without proof in 1867, was finally proved through the ingenuity of the prominent Hungarian mathematician Paul Turán (1910–1976). By the early twentieth century, Chinese mathematical schools had marked out their own territory, specializing in standard areas of mathematics such as analytic function theory. Despite the difficulties of war, revolution, and a period of isolation during the 1960s, transmission of mathematical literature between China and the West continued and greatly expanded through exchanges of students and faculty from the 1980s onward. Kazdan (1986) gives an interesting snapshot of the situation in China at the beginning of this period of expansion.

22.3. THE CHINESE NUMBER SYSTEM

In contrast to the Egyptians, who computed with ink on papyrus, the ancient Chinese, starting in the time of the Shang Dynasty, used rods representing numerals to carry out computations.

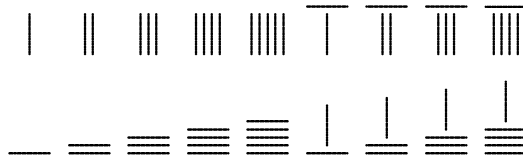


Figure 22.1. The Shang numerals.

Chinese documents from the second century BCE mention the use of counting rods, and a set of such rods from the first century BCE was discovered in 1970. The rods can be arranged to form the Shang numerals (Fig. 22.1) and thereby represent decimal digits. They were used in conjunction with a counting board, which is a board ruled into squares so that each column (or row, depending on the direction of writing) represents a particular item. In pure computations, the successive rows in the board indexed powers of 10. These rods could be stacked to represent any digit from 1 to 9. Since they were placed on a board in rows and columns, the empty places are logically equivalent to a use of 0, but not psychologically equivalent. The use of a circle for zero in China is not found before the thirteenth century. On the other hand, according to Lam and Ang (1987, p. 102), the concept of negative numbers (*fu*), represented by black rods instead of the usual red ones for positive numbers (*cheng*), was also present as early as the fourth century BCE.

It is difficult to distinguish between, say, 22 (|||) and 4 (||||) if the rods are placed too close together. To avoid that difficulty, the Chinese rotated the rods in alternate rows through a right angle, in effect using a positional system based on 100 rather than 10. Since this book is being published in a language that is read from left to right, then from top to bottom, we shall alternate columns rather than rows. In our exposition of the system the number 22 becomes =|| and 4 remains |||. The Shang numerals are shown in Fig. 22.1, the top row being used to represent digits multiplied by an even power of 10 and the bottom row representing digits multiplied by an odd power of 10.

22.3.1. Fractions and Roots

The *Sun Zi Suan Jing* gives a procedure for reducing fractions that is equivalent to the familiar Euclidean algorithm for finding the greatest common divisor of two integers. The rule is to subtract the smaller number from the larger until the difference is smaller than the originally smaller number. Then begin subtracting the difference from the smaller number. Continue this procedure until two equal numbers are obtained. That number can then be divided out of both numerator and denominator.

With this procedure for reducing fractions to lowest terms, a complete and simple theory of computation with fractions is feasible. Such a theory is given in the *Sun Zi Suan Jing*, including the standard procedure for converting a mixed number to an improper fraction and the procedures for adding, subtracting, multiplying, and dividing fractions. Thus, the Chinese had complete control over the system of rational numbers, including, as we shall see below, the negative rational numbers.

At an early date the Chinese dealt with roots of integers, numbers like $\sqrt{355}$, which we now know to be irrational; and they found mixed numbers as approximations when the integer is not a perfect square. In the case of $\sqrt{355}$, the approximation would have been given as $18\frac{31}{36}$. (The denominator is always twice the integer part, as a result of the

approximation used. As with the Hindus and others, the basic principle is that $\sqrt{a+r} \approx \sqrt{a}(1+r/2a) = \sqrt{a} + r/2\sqrt{a}$.

22.4. ALGEBRA

Sooner or later, constantly solving problems of more and more complexity in order to find unknown quantities leads to the systematization of ways of imagining operations performed on a “generic” number (unknown). When the point arises at which an unknown or unspecified number is described by some of its properties rather than explicitly named, we may say that algebra has arisen. There is a kind of twilight zone between arithmetic and algebra, in which certain problems are solved imaginatively without using symbols for unknowns, but later are seen to be easily solvable by the systematic methods of algebra. An example of such a problem is the one from the Bakshali manuscript discussed in Chapter 20 asking for the relative prices of draft horses, thoroughbred horses, and camels.

A good example from China is Problem 15 of Chapter 3 of the *Sun Zi Suan Jing*, which asks how many carts and how many people are involved, given that there are two empty carts (and all the others are full) when people are assigned three to a cart, but nine people have to walk if only two are placed in each cart. We would naturally make this a problem in two linear equations in two unknowns: If x is the number of people and y the number of carts, then

$$x = 3(y - 2),$$

$$x = 2y + 9.$$

However, that would be using algebra, and Sun Zi does not quite do that in this case. His solution is as follows:

Put down 2 carts, multiply by 3 to give 6, add 9, which is the number of persons who have to walk, to obtain 15 carts. To find the number of persons, multiply the number of carts by 2 and add 9, which is the number of persons who have to walk.

Probably the reasoning in the first sentence here is pictorial. Imagine each cart filled with three people. When loaded in this way, the carts would accommodate all the “real” people in the problem, plus six “fictitious” people, since we are given that two carts would be empty if the others each carried three people. Let us imagine, then, that six fictitious people are added to the passengers, one in each of six carts, each of which therefore contains two real people and one fictitious person, while each of the others contains three real people. Now imagine one person removed from each cart, preferably a fictitious person if possible. The number of people removed would obviously be equal to the number of carts. The six fictitious people would then be removed, along with the nine real people who have to walk when there are only two people in each cart. It follows that there must be 15 carts. Finding the number of people (39) is straightforward once the number of carts is known.

The nature of divisibility for integers is also studied in the *Sun Zi Suan Jing*, which contains the essence of the result still known today as the Chinese remainder theorem. The problem asks for a number that leaves a remainder of 2 when divided by 3, a remainder of 3 when divided by 5, and a remainder of 2 when divided by 7. The fact that any number

of such congruences can be solved simultaneously if the divisors are all pairwise relatively prime is the content of what we know now as the Chinese remainder theorem. According to Dickson (1920, p. 57), this name arose when the mathematically literate British missionary Alexander Wylie (1815–1887) wrote an article on it in the English-language newspaper *North China Herald* in 1852. By that time the result was already known in Europe, having been discovered by Gauss and published in his *Disquisitiones arithmeticae* (Art. 36) in 1836.

Sun Zi's answer to this problem shows that he knew a general method of proceeding. He says, "Since the remainder on division by 3 is 2, take 140. The remainder on division by 5 is 3, so take 63. The remainder on division by 7 is 2, so take 30. Add these numbers, getting 233. From this subtract 210, getting the answer as 23." In other words, he took the smallest multiple of $5 \cdot 7$ that leaves a remainder of 2 when divided by 3, then the smallest multiple of $3 \cdot 7$ that leaves a remainder of 3 when divided by 5, and then the smallest multiple of $3 \cdot 5$ that leaves a remainder of 2 when divided by 7. The sum of these numbers was bound to satisfy all three congruences, and then he could add or subtract an arbitrary multiple of $3 \cdot 5 \cdot 7$.

22.5. CONTENTS OF THE *JIU ZHANG SUAN SHU*

This classic work assumes that the methods of calculation explained in the *Sun Zi Suan Jing* are known and applies them to problems very similar to those discussed in the Rhind papyrus. In fact, Problems 5, 7, 10, and 15 from the Chapter 1 of the *Jiu Zhang Suan Shu* are reprinted at the beginning of Chapter 2 of the *Sun Zi Suan Jing*. As its title implies, the book is divided into nine chapters. These nine chapters contain a total of 246 problems. The first eight of these chapters discuss calculation and problems that we would now solve using linear algebra. The last chapter is a study of right triangles and will be discussed below. First, we summarize the contents of some of the earlier parts.

The first chapter, whose title is "Rectangular Fields," discusses how to express the areas of fields given their sides. Problem 1, for example, asks for the area of a rectangular field that is 15 *bu* by 16 *bu*.³ The answer, we see immediately, is 240 "square *bu*." However, the Chinese original does not distinguish between linear and square units. The answer is given as "1 *mu*." The *Sun Zi Suan Jing* explains that as a unit of *length*, 1 *mu* equals 240 *bu*. This ambiguity is puzzling, since a *mu* is both a length equal to 240 *bu* and the area of a rectangle whose dimensions are 1 *bu* by 240 *bu*. It would seem more natural for us if 1 *mu* of area were represented by a square of side 1 *mu*. If these units were described consistently, a square of side 1 linear *mu* would have an area equal to 240 "areal" *mu*. That there really is such a consistency appears in Problems 3 and 4, in which the sides are given in *li*. Since 1 *li* equals 300 *bu* (that is, 1.25 *mu*), to convert the area into *mu* one must multiply the lengths of the sides in *li* and then multiply by $1.25^2 \cdot 240 = 375$. Thus, one gets first "square *mu*" in the sense that we would understand it, and this numerical value for the area is then multiplied by the standard unit shape of 1×240 *bu*. The instructions say to multiply by precisely that number, and the answer is represented as a rectangle 1 *bu* by 375 *bu*.

Chapter 2 ("Millet and Rice") of the *Jiu Zhang Suan Shu* contains problems very similar to the *pesu* problems from the Rhind papyrus. The proportions of millet and various kinds

³One *bu* is 600,000 *hu*, a *hu* being the diameter of a silk thread as it emerges from a silkworm. From other sources, it appears that 1 *bu* is about 1.5 meters.

of rice and other grains are given as empirical data at the beginning of the chapter. Problems of the sort studied in this chapter occur frequently in all commercial transactions in all times. In the United States, for example, a concept analogous to *pesu* is the *unit price* (the number of dollars the merchant will obtain by selling 1 unit of the commodity in question). This number is frequently printed on the shelves of grocery stores to enable shoppers to compare the relative cost of purchasing different brands. Thus, the practicality of this kind of calculation is obvious. The 46 problems in Chapter 2, and also the 20 problems in Chapter 3 (“Proportional Distribution”) of the *Jiu Zhang Suan Shu* are of this type, including some extensions of the Rule of Three. For example, Problem 20 of Chapter 3 asks for the interest due on a loan of 750 *qian* repaid after 9 days if a loan of 1000 *qian* earns 30 *qian* interest each month (a month being 30 days). The result is obtained by forming the product 750 *qian* times 30 *qian* times 9 days and then dividing by the product 1000 *qian* times 30 days, yielding $6\frac{3}{4}$ *qian*. Here the product of the monthly interest on a loan of 1 *qian* and the number of days the loan is outstanding, divided by 30, forms the analog of the *pesu* for the loan; it is the number of *qian* of interest produced by each *qian* loaned.

Chapter 6 (“Fair Transportation”) is concerned with the very important problem of fair allocation of the burdens of citizenship. The Chinese idea of fairness, like that in many other places, including modern America, involves direct proportion. For example, Problem 1 considers a case of collecting taxes in a given location from four counties lying at different distances from the collection center and having different numbers of households. To solve this problem, a constant of proportionality is assigned to each county equal to the number of its households divided by its distance from the collection center. The amount of tax (in millet) each county is to provide is its constant divided by the sum of all the constants of proportionality and multiplied by the total amount of tax to be collected. The number of carts (of a total prescribed number) to be provided by each county is determined the same way. The data in the problem are as follows.

County	Number of Households	Distance to Collection Center
A	10,000	8 days
B	9,500	10 days
C	12,350	13 days
D	12,200	20 days

A total of 250,000 *hu* of millet were to be collected as tax, using 10,000 carts. The proportional parts for the four counties were therefore $10,000 \div 8 = 1250$, $9500 \div 10 = 950$, $12,350 \div 13 = 950$, and $12,200 \div 20 = 610$, which the author reduced to 125, 95, 95, and 61. These numbers total 376. It therefore followed that county A should provide $\frac{125}{376} \cdot 250,000$ *hu*, that is, approximately 83,111.7 *hu* of millet and $\frac{125}{376} \cdot 10,000$, or 3324 carts. The author rounded off the tax to three significant digits, giving it as 83,100 *hu*.

Along with these administrative problems, the 28 problems of Chapter 6 also contain some problems that have acquired an established place in algebra texts throughout the world and will be continue to be worked by students as long as there are teachers to require it. For example, Problem 26 considers a pond used for irrigation and fed by pipes from five different sources. Given that these five canals, each “working” alone, can fill the pond in $\frac{1}{3}$, 1, $2\frac{1}{2}$, 3, and 5 days, the problem asks how long all five “working” together will require to fill it. The author realized that the secret is to add the rates at which the pipes “work” (the reciprocals of the times they require individually to fill the pond) and then take the reciprocal of this sum, and this instruction is given. The answer is $1/(3 + 1 + 2/5 + 1/3 + 1/5) = 15/74$.

22.6. EARLY CHINESE GEOMETRY

Three early Chinese documents contain some geometry, always connected with the computation of areas and volumes. We shall discuss the geometry in them in chronological order.

22.6.1. The *Zhou Bi Suan Jing*

As mentioned above, the earliest Chinese mathematical document still in existence, the *Zhou Bi Suan Jing*, is concerned with astronomy and the applications of mathematics to the study of the heavens. The title refers to the use of the sundial or gnomon in astronomy. This is the physical model that led the Chinese to discover the Pythagorean theorem. Here is a paraphrase of the discussion:

Cut a rectangle whose width is 3 units and whose length is 4 (units) along its diagonal. After drawing a square on this diagonal, cover it with half-rectangles identical to the piece of the original rectangle that lies outside the square, so as to form a square of side 7. [See Fig. 22.2.] Then the four outer half-rectangles, each of width 3 and length 4, equal two of the original rectangles and hence have area 24. When this amount is subtracted from the square of area 49, the difference, which is the area of the square on the diagonal, is seen to be 25. The length of the diagonal is therefore 5.

Although the proof is given only for the easily computable case of the 3–4–5 right triangle, it is obvious that the geometric method is perfectly general, lacking only abstract symbols for unspecified numbers. In our terms, the author has proved that the length of the diagonal of a rectangle whose width is a and whose length is b is the square root of $(a + b)^2 - 2ab$. Note that this form of the theorem is not the “ $a^2 + b^2 = c^2$ ” that we are familiar with. The diagram is shown in Fig. 22.2 for the special case of a 3–4–5 triangle.

According to Li and Du (1987, p. 29), the vertical bar on a sundial was called *gu* in Chinese, and its shadow on the sundial was called *gou*; for that reason the Pythagorean theorem was known as the *gougu* theorem. Cullen (1996, p. 77) says that *gu* means *thigh* and *gou* means *hook*. All authorities agree that the hypotenuse was called *xian* (bowstring),

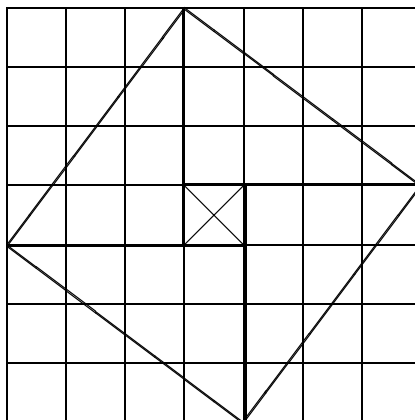


Figure 22.2. Chinese illustration of the Pythagorean theorem.

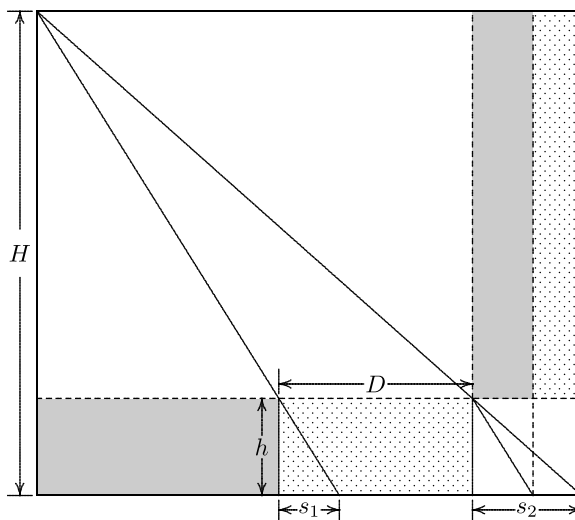


Figure 22.3. The double-difference method of surveying.

which was also Aryabhata's name for it. The *Zhou Bi Suan Jing* says that the Emperor Yu was able to bring order into the realm because he knew how to use this theorem to compute distances. Zhao Shuang credited the Emperor Yu with saving his people from floods and other great calamities, saying that in order to do so he had to survey the shapes of mountains and rivers. Apparently the Emperor had drainage canals dug to channel floods out of the valleys and into the Yangtze and Yellow Rivers.

The third-century commentary on the *Zhou Bi Suan Jing* by Zhao Shuang explains a close variant of the method of surveying discussed in connection with the work of Aryabhata I in Chapter 20. The Chinese variant of the method is illustrated in Fig. 22.3, which assumes that the height H of an inaccessible object is to be determined. To determine H , it is necessary to put two poles of a known height h vertically into the ground in line with the object at a known distance D apart. The height h and the distance D are theoretically arbitrary, but the larger they are, the more accurate the results will be. After the poles are set up, the lengths of the shadows they would cast if the sun were at the inaccessible object are measured as s_1 and s_2 . Thus the lengths s_1 , s_2 , h , and D are all known. A little trigonometry and algebra will show that

$$H = h + \frac{Dh}{s_2 - s_1}.$$

We have given the result as a formula, but as a set of instructions it is very easy to state in words: *The required height is found by multiplying the height of the poles by the distance between them, dividing by the difference of the shadow lengths, and adding the height of the poles.*

This method was expounded in more detail in a commentary on the *Jiu Zhang Suan Shu* written by Liu Hui in 263 CE. This commentary, along with the rest of the material on right triangles in the *Jiu Zhang Suan Shu*, eventually became a separate treatise, the *Hai Dao Suan Jing* (*Sea Island Mathematical Manual*; see Ang and Swetz, 1986). Liu Hui mentioned that this method of surveying could be found in the *Zhou Bi Suan Jing* and called it the *double*

difference method (chong cha). The name apparently arises because the difference $H - h$ is obtained by dividing Dh by the difference $s_2 - s_1$.

We have described the lengths s_1 and s_2 as shadow lengths here because that is the problem used by Zhao Shuang to illustrate the method of surveying. He attempts to calculate the height of the sun, given that at the summer solstice a stake 8 *chi* high casts a shadow 6 *chi* long and that the shadow length decreases by 1 *fen* for every 100 *li* that the stake is moved south, casting no shadow at all when moved 60,000 *li* to the south. This model assumes a flat earth, under which the shadow length is proportional to the distance from the pole to the foot of the perpendicular from the sun to the plane of the earth. Even granting this assumption, as we know, the sun is so distant from the earth that no lengthening or shortening of shadows would be observed. To any observable precision the sun's rays are parallel at all points on the earth's surface. The small change in shadow length that we observe is due to the curvature of the earth. But let us continue, accepting Zhao Shuang's assumptions.

To explain the solution of this problem, we first observe that two pieces of data are irrelevant to the problem. It does not matter how long a shadow is, since only the difference $s_1 - s_2$ occurs in the computational procedure. Likewise, the statement that the shadow disappears at a certain location (which, by the way, lies at an impossible distance away—the earth is not that big!) is irrelevant. The data here are $D = 1000$ *li*, $s_2 - s_1 = 1$ *fen*, $h = 8$ *chi*. One *chi* is about 25 centimeters, one *fen* is about 2.5 cm, and one *li* is 1800 *chi*, that is, about 450 meters. Our first job is to express everything in consistent units, say *li*. Thus, $D = 1000$, $s_2 - s_1 = \frac{1}{180,000}$, and $h = \frac{8}{1800}$.

Because the pole height h is obviously insignificant in comparison with the height of the sun, we can neglect the first term in the formula we gave above, and we write

$$H = \frac{Dh}{s_2 - s_1}.$$

When we insert the appropriate values, we find, as did Zhao Shuang, that the sun is 80,000 *li* high, about 36,000 kilometers. Later Chinese commentators recognized that this figure was inaccurate, and in the eighth century an expedition to survey accurately a north–south line found the actual lengthening of the shadow to be 4 *fen* per thousand *li*. Notice that the statement of the problem seems to reveal careless editing over the years, since two methods of computing the height are implied here. If the two irrelevant pieces of information provided are taken into account, one can immediately use the similar triangles to infer the height of 80,000 *li*. This fact suggests that the original text was modified by later commentators, but that not all the parts that became irrelevant as a result of the modifications were removed.

22.6.2. The *Jiu Zhang Suan Shu*

The *Jiu Zhang Suan Shu* contains all the standard formulas for the areas of squares, rectangles, triangles, and trapezoids, along with the recognition of a relation between the circumference and the area of a circle, which we could interpret as a connection between the one-dimensional π and the two-dimensional π . The geometric formulas given in this treatise are more extensive than those of the Rhind papyrus; for example, there are approximate formulas for the volume of segment of a sphere and the area of a segment of a circle. It is perhaps not fair to compare the two documents, since the Rhind papyrus was written nearly two millennia earlier. The implied value of one-dimensional π , however, is

$\pi = 3$. It is surprising to find this value so late, since it is known that the value 3.15147 had been obtained in China by the first century. According to Li and Du (1987, p. 68), Liu Hui refined it to $3.14 + 64/62500 = 3.141024$ by approximating the area of a 192-sided polygon.⁴ That is, he started with a hexagon and doubled the number of sides five times.

Problems 31 and 32 ask for the area of a circular field of a given diameter and circumference.⁵ The method is to multiply half of the circumference by half of the diameter, which is exactly right in terms of Euclidean geometry; equivalently, the reader is told that one may multiply the two dimensions and divide by 4. In the actual data for problems, the diameter given is exactly one-third of the given circumference; in other words, the value assumed for one-dimensional π is 3. The assumption of that value leads to two other procedures for calculating the area: squaring the diameter, then multiplying by 3 and dividing by 4, or squaring the circumference and dividing by 12. An elaboration of this problem occurs in Problems 37 and 38, in which the area of an annulus (the region outside the smaller of two concentric circles and inside the larger) is given in terms of its width and the circumferences of the two circles.

The authors knew also how to find the volume of a pyramid. Problem 15 of Chapter 5 asks for the volume of a pyramid whose base is a rectangle 5 *chi* by 7 *chi* and whose height is 8 *chi*. The answer is given (correctly) as $93\frac{1}{3}$ (cubic) *chi*. For a frustum of a pyramid having rectangular bases the recipe is to add twice the length of the upper base to the lower base and multiply by the width of the upper base to get one term. A second term is obtained symmetrically as twice the length of the lower base plus the length of the upper base, multiplied by the width of the lower base. These two terms are then added and multiplied by the altitude, after which one divides by 6. If the bases are $a \times b$ and $c \times d$ (the sides of length a and c being parallel) and the height is h , this yields what we would write (correctly) as

$$V = \frac{h}{6} [(2a + c)b + (2c + a)d].$$

This result is more general than the rule given in the Moscow papyrus discussed in Section 7.2 of Chapter 7, which is given for a frustum with square bases.

The Pythagorean Theorem The last of the nine chapters of the *Jiu Zhang Suan Shu* contains 24 problems on the *gougu* theorem. After a few “warm-up” problems in which two of the three sides of a right triangle are given and the third is to be computed, the problems become more complicated. Problem 11, for example, gives a rectangular door whose height exceeds its width by 6 *chi*, 8 *cun* and has a diagonal of 1 *zhang*. One *zhang* is 10 *chi* and 1 *chi* is 10 *cun* (apparently a variant rendering of *fen*). The recipe given is correct: Take half the difference of the height and width, square it, double, subtract from the square of the diagonal, then take the square root of half of the result. That process yields the average of the height and width, and given their semidifference of 3 *chi*, 4 *cun*, one can easily get both the width and the height.

⁴Lam and Ang (1986) give the value as $3.14 + 169/625 = 3.142704$.

⁵All references to problem numbers and nomenclature in this section are based on the article of Lam (1994).

22.6.3. The *Sun Zi Suan Jing*

The *Sun Zi Suan Jing* contains a few problems in measurement that are unusual enough to merit some discussion. An inverse area problem occurs in Problem 20, in which a circle is said to have area 35,000 square *bu*, and its circumference is required. Since the area is taken as one-twelfth of the square of the circumference, the author multiplies by 12, then takes the square root, getting $648\frac{96}{1296}$ *bu*.

PROBLEMS AND QUESTIONS

Mathematical Problems

- 22.1. Compare the pond-filling problem (Problem 26 of Chapter 6) of the *Jiu Zhang Suan Shu* (discussed above) with the following problem from Greenleaf (1876, p. 125): *A cistern has three pipes; the first will fill it in 10 hours, the second in 15 hours, and the third in 16 hours. What time will it take them all to fill it?* Solve this problem. Is there any real difference between the two problems?
- 22.2. What happens to the estimate of the sun's altitude (36,000 km) given by Zhao Shuang if the "corrected" figure for shadow lengthening (4 *fen* per 1000 *li*) is used in place of the figure of 1 *fen* per 1000 *li*?
- 22.3. The *gougu* section of the *Jiu Zhang Suan Shu* contains the following problem: *Under a tree 20 feet high and 3 in circumference there grows a vine, which winds seven times the stem of the tree and just reaches its top. How long is the vine?*
Solve this problem. [Hint: Picture the tree as a cylinder. Imagine it has been cut down and rolled along the ground in the direction perpendicular to its axis in order to unwind the vine onto the ground. What would the trace of the tree and the vine on the ground look like?]

Historical Questions

- 22.4. What uses were claimed for mathematics in the early Chinese classics?
- 22.5. What kinds of problems are studied in the nine chapters of the *Jiu Zhang Suan Shu*?
- 22.6. How is the Pythagorean theorem treated in the *Zhou Bi Suan Jing*?

Questions for Reflection

- 22.7. The fair taxation problem from the *Jiu Zhang Suan Shu* considered above treats distances and population with equal weight. That is, if the population of one county is double that of another, but that county is twice as far from the collection center, the two counties will have exactly the same tax assessment in grain and carts. Will this impose an equal burden on the taxpayers of the two counties? Is there a direct proportionality between distance and population that makes them interchangeable from the point of view of the taxpayers involved? Is the growing of extra grain to pay the tax fairly compensated by a shorter journey?

- 22.8.** The *Jiu Zhang Suan Shu* implies that the diameter of a sphere is proportional to the cube root of its volume. Since this fact is equivalent to saying that the volume is proportional to the cube of the diameter, should we infer that the author knew both proportions? More generally, if an author knows (or has proved) “fact A,” and fact A implies fact B, is it accurate to say that the author knew or proved fact B?
- 22.9.** How is the remainder problem of Sun Zi related to the *kuttaka* method of Brahmagupta. [*Hint:* The statement that x leaves a remainder of (say) 5 when divided by (say) 38 can be interpreted as saying there is an integer y such that $x = 38y + 5$. If you also want x to leave a remainder of (say) 4 when divided by (say) 15, you can write the equation $x = 15z + 4$. Eliminating x , you find $15z + 4 = 38y + 5$, that is, $15z = 38y + 1$. How do you find *all* solutions of this Diophantine equation?]

Later Chinese Algebra and Geometry

We begin our examination of Chinese algebra by taking a brief look at number theory in China. Unlike the Greeks, Chinese mathematicians were not interested in figurate numbers. Still, there was in China an interest in the use of numbers for divination. According to Li and Du (1987, pp. 95–97), the magic square

4	9	2
3	5	7
8	1	6

appears in the treatise *Shushu Jiayi* (*Memoir on Some Traditions of the Mathematical Art*) by the sixth-century mathematician Zhen Luan. In this figure each row, column, and diagonal totals 15. In the early tenth century, during the Song Dynasty, a connection was made between this magic square and a figure called the *Luo-chu-shu* (*book that came out of the River Lo*) found in an appendix to the *Book of Changes*. The *Book of Changes* states that a tortoise crawled out of the River Lo and delivered to the Emperor Yu the diagram in Fig. 23.1. Because of this connection, the diagram came to be called the *Luo-shu* (*Luo book*). The purely numerical aspects of the magic square are enhanced by representing the even (female, *ying*) numbers as solid disks and the odd (male, *yang*) numbers as open circles. Like so much of number theory, the theory of magic squares has continued to attract attention from specialists, despite being devoid of applications. The interest has come from combinatoricists, for whom Latin squares¹ are a topic of continuing research.

23.1. ALGEBRA

The development of algebra in China began early and continued for many centuries. The aim was to find numerical approximate solutions to equations, and the Chinese mathematicians were not intimidated by equations of high degree.

¹A Latin square is a square array of symbols in which each symbol occurs precisely once in each row and precisely once in each column.

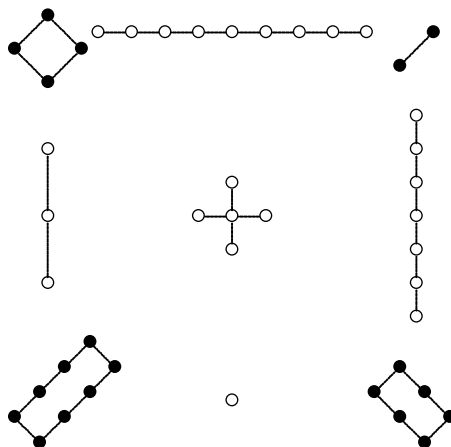


Figure 23.1. The *Luo-shu*.

23.1.1. Systems of Linear Equations

Systems of linear equations occur in the *Jiu Zhang Suan Shu* (Mikami, 1913, pp. 18–22; Li and Du, 1987, pp. 46–49). Here is one example of the technique.

There are three kinds [of wheat]. The grains contained in two, three and four bundles, respectively, of these three classes [of wheat], are not sufficient to make a whole measure. If, however, we add to them one bundle of the second, third, and first classes, respectively, then the grains would become one full measure in each case. How many measures of grain does then each one bundle of the different classes contain?

The following counting-board arrangement is given for this problem.

1		2	1st class
	3	1	2nd class
4	1		3rd class
1	1	1	measures

Here the three columns of numbers from right to left represent the three samples of wheat. Thus the right-hand column represents 2 bundles of the first class of wheat, to which one bundle of the second class has been added. The bottom row gives the result in each case: 1 measure of wheat. The word problem might be clearer if the final result is thought of as the result of threshing the raw wheat to produce pure grain. Without seriously distorting the procedure followed by the author, we can write down this counting board as a matrix and solve the resulting system of three equations in three unknowns. The author gives the solution: A bundle of the first type of wheat contains $\frac{9}{25}$ measure, a bundle of the second contains $\frac{7}{25}$ measure, and a bundle of the third contains $\frac{4}{25}$ measure.

23.1.2. Quadratic Equations

The last chapter of the *Jiu Zhang Suan Shu*, which involves right triangles, contains problems that lead to linear and quadratic equations. For example (Mikami, 1913, p. 24), there are

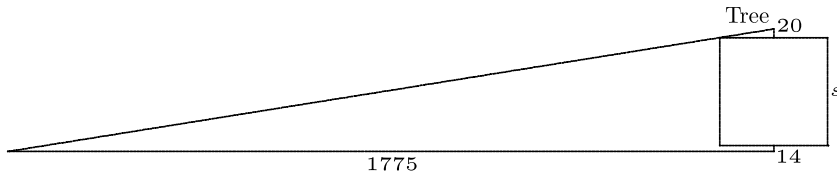


Figure 23.2. A quadratic equation problem from the *Jiu Zhang Suan Shu*.

several problems involving a town enclosed by a square wall with a gate in the center of each side. In some cases the problem asks at what distance (x) from the south gate a tree a given distance east of the east gate will first be visible. The data are the side s of the square and the distance d of the tree from the gate. For that kind of data, the problem is the linear equation $2x/s = s/(2d)$. When the side of the town (s) is the unknown, a quadratic equation results. In one case, it is asserted that the tree is 20 paces north of the north gate and is just visible to a person who walks 14 paces south of the south gate, then 1775 paces west. This problem proposes the quadratic equation $s^2 + 34s = 71000$ to be solved for the unknown side s . (See Fig. 23.2, which is drawn to scale to show how unrealistic the problem really is.) Since the Chinese technique of solving equations numerically is practically independent of degree, we shall not bother to discuss the techniques for solving quadratic equations separately.

23.1.3. Cubic Equations

Cubic equations first appear in Chinese mathematics (Li and Du, 1987, p. 100; Mikami, 1913, p. 53) in the seventh-century work *Xugu Suanjing* (*Continuation of Ancient Mathematics*) by Wang Xiaotong (ca. 580–ca. 640). This work contains some intricate problems associated with right triangles. For example, compute the length of a leg of a right triangle given that the product of the other leg and the hypotenuse is $1337 \frac{1}{20}$ and the difference between the hypotenuse and the leg is $1 \frac{1}{10}$. (Mikami gives $\frac{1}{10}$ as the difference, which is incompatible with the answer given by Wang Xiaotong. I do not know if the mistake is due to Mikami or is in the original.) This problem is easy to state for a general product P and difference D . Wang Xiaotong gives a general description of the result of eliminating the hypotenuse and the other leg that amounts to the equation

$$x^3 + \frac{5D}{2}x^2 + 2D^2x = \frac{P^2}{2D} - \frac{D^3}{2}.$$

In the present case (using the corrected data) the equation is

$$x^3 + \frac{11}{4}x^2 + \frac{121}{50}x - 812591 \frac{59}{125} = 0.$$

Wang Xiaotong then gives the root (correctly) as $92 \frac{2}{5}$ “according to the rule of the cubic root extraction.” Li and Du (1987, pp. 118–119) report that the eleventh-century mathematician Jia Xian developed a method for extracting the cube root that generalizes from the case of the equation $x^3 = N$ to the general cubic equation, and even to an equation of arbitrarily high degree, at least in theory, as we shall now see.

23.1.4. A Digression on the Numerical Solution of Equations

The Chinese mathematicians of 800 years ago invented a method of finding numerical approximations of a root of an equation, similar to a method that was rediscovered independently in the nineteenth century in Europe and is commonly called *Horner's method*, in honor of the British school teacher William Horner (1786–1837).² The first appearance of the method is in the work of the thirteenth-century mathematician Qin Jiushao, who applied it in his 1247 treatise *Sushu Jiu Zhang* (*Arithmetic in Nine Chapters*, not to be confused with the *Jiu Zhang Suan Shu*).

We illustrate with the case of the cubic equation. Suppose in attempting to solve the cubic equation $px^3 + qx^2 + rx + s = 0$ we have found an initial approximation a for the root. (Typically, this is done by getting the first digit or the integer part of the root.) We then “reduce” the equation by setting $x = y + a$ and rewriting it. What will the coefficients be when the equation is written in terms of y ? The answer is immediate; the new equation is

$$\begin{aligned} & py^3 + 3pay^2 + 3pa^2y + pa^3 \\ & + \quad qy^2 + 2qay + qa^2 \\ & \quad \quad + \quad ry + ra \\ & \quad \quad \quad + \quad s = 0. \end{aligned}$$

We see that we need to make the following conversion of the coefficients:

$$\begin{array}{r} p \qquad \qquad p \\ q \qquad \qquad 3pa + q \\ r \longrightarrow 3pa^2 + 2qa + r \\ s \qquad pa^3 + qa^2 + ra + s \end{array}$$

Here is a method of making this reduction that is well adapted for use on a counting board.

1. *Step 1:* By inspection, find a first approximation to a root. (Simply evaluate the polynomial at, say 1, 10, 100, and so on, finding out where it changes sign from negative to positive or vice versa. If it is negative at 10 and positive at 100, for example, then evaluate it at 20, 30, 40, and so on, until you find again where it changes sign. If it is negative at 30 and positive at 40, for example, then you can take 30 as the first approximation.)
2. *Step 2:* Lay out a template in the form of a 4×5 matrix (for cubic equations), in which (1) all the entries in the top row are the same, namely the leading coefficient,

²Besides being known to the Chinese mathematicians 600 years before Horner, this procedure was used by Sharaf al-Tusi (ca. 1135–1213) and was discovered by the Italian mathematician Paolo Ruffini (1765–1822) a few years before Horner published it. In fairness to Horner, it must be said that he applied the method not only to polynomials, but to infinite series representations. To him it was a theorem in calculus, not algebra.

(2) the first column is the coefficients, in order, and (3) the lower right triangle consists of zeros. Thus, we get the matrix

$$\begin{matrix} p & p & p & p & p \\ q & & & & 0 \\ r & & & 0 & 0 \\ s & & 0 & 0 & 0 \end{matrix}$$

3. *Step 3:* Fill in the rest of the entries working from left to right and top to bottom (in either order). In each unoccupied place, put the product of the current approximation a and the entry directly above, plus the entry directly to the left. Thus, we could begin by filling in either the second row or the second column:

$$\begin{matrix} p & p & p & p & p & & p & p & p & p & p \\ q & pa + q & 2pa + q & 3pa + q & 0 & & q & pa + q & & & 0 \\ r & & & 0 & 0 & \text{or} & r & pa^2 + qa + r & & 0 & 0 \\ s & & 0 & 0 & 0 & & s & pa^3 + qa^2 + ra + s & 0 & 0 & 0 \end{matrix}$$

When we finish, we have the following matrix, and the coefficients are read, in order, off the diagonal running from the top right to the bottom of the second column:

$$\begin{matrix} p & p & p & p & p \\ q & pa + q & 2pa + q & 3pa + q & 0 \\ r & pa^2 + qa + r & 3pa^2 + 2qa + r & 0 & 0 \\ s & pa^3 + qa^2 + ra + s & 0 & 0 & 0 \end{matrix}$$

Thus, as we see, the new equation for y is

$$py^3 + (3pa + q)y^2 + (3pa^2 + 2qa + r)y + (pa^3 + qa^2 + ra + s) = 0.$$

The zeros used to form the template for the algorithm have a very important use when the solution is being obtained one digit at a time. It is useful to have a solution between 0 and 10 at each stage, and one way to ensure that, after the integer part of the solution (say a) has been obtained, is to multiply the fractional part y by 10. Then one need only seek the integer part of $z = 10y$. Since z satisfies

$$pz^3 + 10(3pa + q)z^2 + 100(3pa^2 + 2qa + r)z + 1000(pa^3 + qa^2 + ra + s) = 0,$$

and the entries in the matrix will often be integers, one can simply adjoin the zeros to the coefficients when forming the new equation, which necessarily has a solution between 0 and 10.

Wang Xiaotong's reference to the use of cube root extraction for solving his equation seems to suggest that this method was known as early as the seventh century. The earliest recorded instance of it, however, seems to be in the treatise of Qin Jiushao, who illustrated it by solving the quartic equation

$$-x^4 + 763200x^2 - 40642560000 = 0.$$

The method of solution gives proof that the Chinese did not think in terms of a quadratic formula. If they had, this equation would have been solved for x^2 using that formula and then x could have been found by taking the square root of any positive root. But Qin Jiushao applied the fourth-degree analog of the method described above to get the solution $x = 840$. (He missed the smaller solution $x = 240$.)

The method needs to be illustrated with an example. Let us take the equation $0.027x^3 - 3.3x - 20 = 0$. When $x = 10$, the left-hand side equals -26 , and when $x = 20$, it equals 130, so we take $a = 10$ as a first approximation. We then fill out the "transition" matrix:

0.027	0.027	0.027	0.027	0.027		0.027	0.027	0.027	0.027	0.027	
0				0	→	0	0.27	0.54	0.81	0	→
-3.3			0	0		-3.3			0	0	
-20		0	0	0		-20		0	0	0	
	0.027	0.027	0.027	0.027	0.027		0.027	0.027	0.027	0.027	0.027
→	0	0.27	0.54	0.81	0	→	0	0.27	0.54	0.81	0
	-3.3	-0.6	4.8	0	0		-3.3	-0.6	4.8	0	0
	-20		0	0	0		-20	-26	0	0	0

The next term y in our approximation to the roots therefore satisfies the equation $0.027y^3 + 0.81y^2 + 4.8y - 26 = 0$, and it is guaranteed to be between 0 and 10, since $x = y + 10$ was found to lie between 10 and 20. In fact, when $y = 3$, the left-hand side is -3.581 , and when $y = 4$, it is 3.088, so that the next digit is 3. We now repeat the process.

0.027	0.027	0.027	0.027	0.027		0.027	0.027	0.027	0.027	0.027	
0.81				0	→	0.81	0.891			0	→
4.8			0	0		4.8	7.473		0	0	
-26		0	0	0		-26	-3.581	0	0	0	
		0.027	0.027	0.027	0.027	0.027		0.027			
	→	0.81	0.891	0.972		0	→				
		4.8	7.473	10.399	0	0					
		-26	-3.581	0	0	0					
					0.027	0.027	0.027	0.027	0.027		
					→	0.81	0.891	0.972	1.053	0	
						4.8	7.473	10.389	0	0	
						-26	-3.581	0	0	0	

We did the operations column-wise this time, just to show that it makes no difference whether we do it by rows or by columns. Since the next digit will be to the right of the decimal point, it makes sense to multiply the equation by 1000 to get a one-digit solution between 0 and 10. If the next correction is w , it will lie between 0 and 1, and $10w$ will lie between 0 and 10. If $z = 10w$, then $0.027z^3 + 10(1.053)z^2 + 100(10.389)z - 1000(3.581) = 0$, as can be seen by multiplying the equation satisfied by w by 1000, then substituting z for $10w$, z^2 for $100w^2$, and z^3 for $1000w^3$. We thus have $0.027z^3 + 10.53z^2 + 1038.9z - 3581 = 0$, and z is guaranteed to be between 0 and 10. If we had not replaced w by z , we would have had to deal with fractions in making our guesses equal to 0.1, 0.2, 0.3, and so on. Once again, we find that when $z = 3$, the left-hand side is -368.801 , and when $z = 4$, it is 744.808 . Thus, the next digit is also 3. We now have the approximation $x = 13.3$. Continuing the process would reveal that $x = 13.333 \dots = 13\frac{1}{3}$.

A word of explanation is needed about the lower triangle of zeros in this method. They can be useful in getting the successive digits to the right of the decimal point if the coefficients of the original equation are all integers. Then, the equation for the next digit can be written down directly by adjoining these zeros to the coefficients one would otherwise read off.

The efficiency of this method in finding approximate roots allowed the Chinese to attack equations involving large coefficients and high degrees. Qin Jiushao (Libbrecht, 1973, pp. 134–136) considered the following problem: *Three li north of the wall of a circular town there is a tree. A traveler walking east from the southern gate of the town first sees the tree after walking 9 li. What are the diameter and circumference of the town?*

This problem appears to be deliberately concocted so as to lead to an equation of high degree. (The diameter of the town could surely be measured directly from inside, so that it is highly unlikely that anyone would ever need to solve such a problem for a practical purpose.) Representing the diameter of the town as x^2 , Qin Jiushao obtained the equation³

$$x^{10} + 15x^8 + 72x^6 - 864x^4 - 11664x^2 - 34992 = 0.$$

One has to be very unlucky to get such a high-degree equation. Even a simplistic approach using similar triangles leads only to a quartic equation. It is easy to see (Fig. 23.2) that if the diameter of the town rather than its square root is taken as the unknown, and the radius is drawn to the point of tangency, trigonometry will yield a quartic equation. If the radius is taken as the unknown, the similar right triangles in Fig. 23.3 lead to the quartic equation $4r^4 + 12r^3 + 9r^2 - 486r - 729 = 0$, and since this equation has $r = -\frac{3}{2}$ as a solution, we can divide the left-hand side by $2r + 3$, getting the cubic equation $2r^3 + 3r^2 - 243 = 0$. But, of course, the object of this game was probably to practice the art of algebra, not to get the simplest possible equation, no matter how virtuous it may seem to do so in other contexts. In any case, the historian's job is not that of a commentator trying to improve a text. It is to try to understand what the original author was thinking. Probably the elevated degree is the result of having to circumvent the use of similar triangles by relying entirely on the Pythagorean theorem. (Even with that assumption, however, it is quite easy to get a quartic equation for the diameter.)

³Even mathematicians working within the Chinese tradition seem to have been puzzled by the needless elevation of the degree of the equation (Libbrecht, 1973, p. 136).

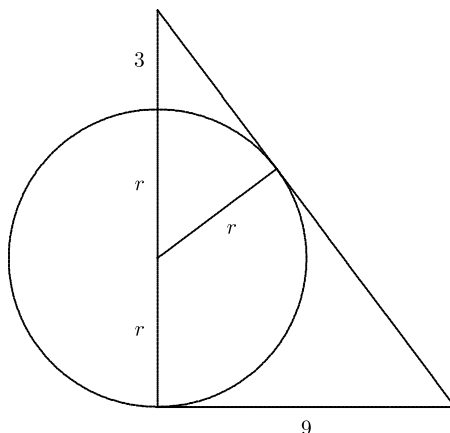


Figure 23.3. A quartic equation problem.

23.2. LATER CHINESE GEOMETRY

Chinese mathematics was greatly enriched from the third through the sixth centuries by a series of brilliant geometers, whose achievements deserve to be remembered alongside those of Euclid, Archimedes, and Apollonius. We have space to discuss only three of these.

23.2.1. Liu Hui

We begin with the third-century mathematician Liu Hui (ca. 220–ca. 280), author of the *Hai Dao Suan Jing* mentioned in the previous chapter. Liu Hui had a remarkable ability to visualize figures in three dimensions. In his commentary on the *Jiu Zhang Suan Shu* he asserted that the circumference of a circle of diameter 100 is 314. In solid geometry, he provided dissections of many geometric figures into pieces that could be reassembled to demonstrate their relative sizes beyond any doubt. As a result, real confidence could be placed in the measurement formulas that he provided. He gave correct procedures, based on such dissections, for finding the volumes enclosed by many different kinds of polyhedra. But his greatest achievement is his work on the volume of the sphere.

The *Jiu Zhang Suan Shu* made what appears to be a very reasonable claim: that the ratio of the volume enclosed by a sphere to the volume enclosed by the circumscribed cylinder can be obtained by slicing the sphere and cylinder along the axis of the cylinder and taking the ratio of the area enclosed by the circular cross section of the sphere to the area enclosed by the square cross section of the cylinder. In other words, it would seem that the ratio is $\pi : 4$. This conjecture seems plausible, since every such section produces exactly the same figure. It fails, however, because of the principle behind Guldin's theorem: The volume of a solid of revolution equals the area revolved about the axis times the distance traveled by the *centroid* of the area. The half of the square that is being revolved to generate the cylinder has a centroid that is farther away from the axis than the centroid of the semicircle inside it, whose revolution produces the sphere; hence when the two areas are multiplied by the two distances, the ratio gets changed. When a circle inscribed in a square is rotated, the ratio of the volumes generated is $2 : 3$, while that of the original areas is $\pi : 4$. Liu Hui noticed that

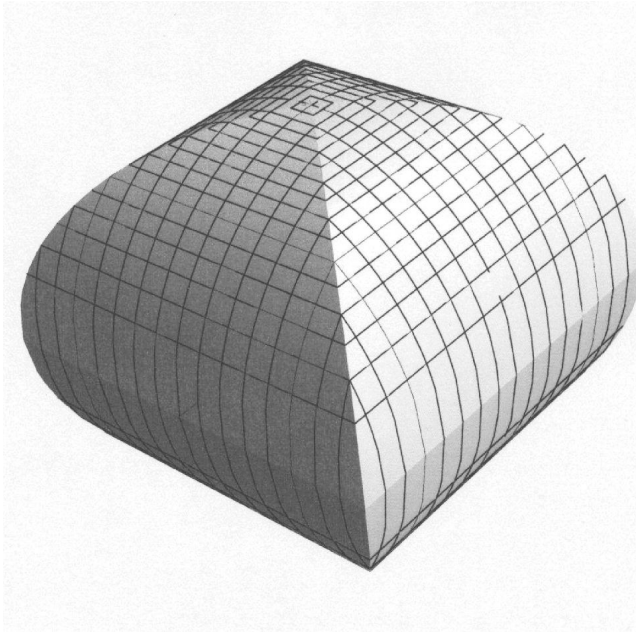


Figure 23.4. The double square umbrella.

the sections of the figure parallel to the base of the cylinder do not all have the same ratios. The sections of the cylinder are all disks of the same size, while the sections of the sphere shrink as the section moves from the equator to the poles. From that observation, he could see that one could not expect the ratio of the cylinder to the inscribed sphere to be the same as that of the square to the inscribed circle.

He also formed a solid by intersecting two cylinders circumscribed about the sphere whose axes are at right angles to each other, thus producing a figure he called a *double square umbrella*, which is now known as a *bicylinder* or *Steinmetz solid*⁴ (see Hogendijk, 2002). A representation of the double square umbrella, generated using *Mathematica* graphics, is shown in Fig. 23.4. Its volume *does* have the same ratio to the sphere that the square has to its inscribed circle, that is, $4 : \pi$. This proportionality between the double square umbrella and the sphere is easy to see intuitively, since every horizontal slice of this figure by a plane parallel to the plane of the axes of the two circumscribed cylinders intersects the double square umbrella in a square and intersects the sphere in the circle inscribed in that square. Liu Hui inferred that the volume enclosed by the double umbrella would have this ratio to the volume enclosed by the sphere. This inference is correct and is an example of what is called *Cavalieri's principle*: *Two solids such that the section of one by each horizontal plane bears a fixed ratio to the section of the other by the same plane have volumes in that same ratio*. This principle had been used by Archimedes five centuries earlier. In the introduction to his *Method*, Archimedes used this very example and asserted (correctly) that the volume of the bicylinder is two-thirds of the volume of the cube in which they are

⁴Named after the German-American mathematician and electrical engineer Charles Proteus Steinmetz (1865–1923).

inscribed.⁵ But Liu Hui's use of it (see Lam and Shen, 1985) was obviously independent of Archimedes. It amounts to a limiting case of the dissections that Liu Hui did so well. The solid is cut into *infinitely thin* slices, each of which is then dissected and reassembled as the corresponding section of a different solid. This realization was a major step toward an accurate measurement of the volume of a sphere. Unfortunately, it was not granted to Liu Hui to complete the journey. He maintained a consistent agnosticism on the problem of computing the volume of a sphere, saying, "Not daring to guess, I wait for a capable man to solve it."

23.2.2. Zu Chongzhi

That "capable man" required a few centuries to appear, and he turned out to be two men. "He" was Zu Chongzhi (429–500) and his son Zu Geng (450–520). Zu Chongzhi was a very capable geometer and astronomer who said that if the diameter of a circle is 1, then the circumference lies between 3.1415926 and 3.1415927. From these bounds, probably using the Chinese version of the Euclidean algorithm, the method of mutual subtraction, he stated that the circumference of a circle of diameter 7 is (approximately) 22 and that of a circle of diameter 113 is (approximately) 355.⁶ These estimates are very good, far too good to be the result of any inspired or hopeful guess. Of course, we don't have to imagine that Zu Chongzhi actually *drew* the polygons. It suffices to know how to compute the perimeter, and that is a simple recursive process: If s_n is the length of the side of a polygon of n sides inscribed in a circle of unit radius, then

$$s_{2n}^2 = 2 - \sqrt{4 - s_n^2}.$$

Hence each doubling of the number of sides makes it necessary to compute a square root, and the approximation of these square roots must be carried out to many decimal places in order to get enough guard digits to keep the errors from accumulating when you multiply this length by the number of sides. In principle, however, given enough patience, one could compute any number of digits of π this way.

One of Zu Chongzhi's outstanding achievements, in collaboration with his son Zu Geng, was finding the volume enclosed by Liu Hui's double square umbrella. As Fu (1991) points out, this volume was not approachable by the direct method of dissection and recombination that Liu Hui had used so successfully.⁷ An indirect approach was needed. The trick turned out to be to enclose the double square umbrella in a cube and look at the volume inside the cube and outside the double square umbrella. Suppose that the sphere has radius R . The double square umbrella can then be enclosed in a cube of side $2R$. Consider a horizontal section of the enclosing cube at height h above the middle plane of that cube. In the double umbrella this section is a square of side $2\sqrt{R^2 - h^2}$ and area $4(R^2 - h^2)$, as shown in

⁵Hogendijk (2002) argues that Archimedes also knew the surface area of the bicylinder.

⁶The approximation $\pi \approx \frac{22}{7}$ was given earlier by He Chengtian (370–447), and of course much earlier by Archimedes. A more sophisticated approach by Zhao Youqin (b. 1271) that gives $\frac{355}{113}$ was discussed by Volkov (1997).

⁷Lam and Shen (1985, p. 223), however, say that Liu Hui *did* consider the idea of setting the double umbrella inside the cube and trying to find the volume between the two. Of course, that volume also is not accessible through direct, finite dissection.

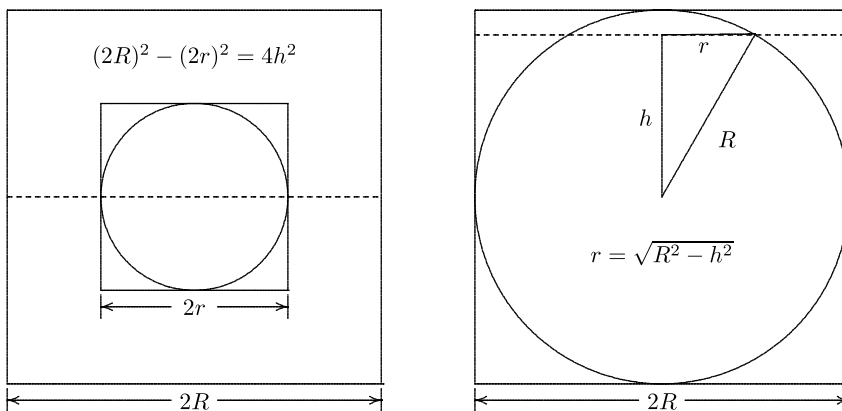


Figure 23.5. Sections of the cube, double square umbrella, and sphere. Left: Horizontal section at height h above the midplane. Right: Vertical section through the center parallel to a side of the cube.

Fig. 23.5. Therefore the area of the section outside the double umbrella and inside the cube is $4h^2$.

It was no small achievement to look at the region in question. It was an even keener insight on the part of the family Zu to realize that this cross-sectional area is equal to the area of the cross section of an upside-down pyramid with a square base of side $2R$ and height R . Hence *the volume of the portion of the cube outside the double umbrella in the upper half of the cube equals the volume of a pyramid with square base of side $2R$ and height R* . But thanks to earlier work contained in Liu Hui's commentaries on the *Jiu Zhang Suan Shu*, Zu Chongzhi knew that this volume was $(4R^3)/3$. It therefore follows, after doubling to include the portion below the middle plane, that the region inside the cube but outside the double umbrella has volume $(8R^3)/3$, and hence that the double umbrella itself has volume $8R^3 - (8R^3)/3 = (16R^3)/3$.

Since, as Liu Hui had shown, the volume of the sphere is $\pi/4$ times the volume of the double square umbrella, it follows that the sphere has volume $(\pi/4) \cdot (16R^3)/3$, or $(4\pi R^3)/3$.

PROBLEMS AND QUESTIONS

Mathematical Problems

- 23.1. Verify the solution of the problem involving three bundles of wheat, for which the solution was given above.
- 23.2. Use the method of the text to get the next two digits of an approximation to a root of the equation $32x^3 - 24x^2 - 60x + 7 = 0$, given that there is a root between 1 and 2. In other words, use $a = 1$ as a first approximation. (Remember, since you are crossing the decimal point, to carry along the extra zeros each time, as was done above.)
- 23.3. Find all the solutions of the cubic equation $2r^3 + 3r^2 = 243$ without doing any numerical approximation. [*Hint*: If there is a rational solution $r = m/n$, then m must divide 243 and n must divide 2.]

Historical Questions

- 23.4. How did Liu Hui demonstrate his geometric theorems?
- 23.5. What kind of algebraic problems did the Chinese solve that were different from those we have discussed from other cultures?
- 23.6. Why were the Chinese mathematicians undeterred by the prospect of solving equations of degree 4 and higher?

Questions for Reflection

- 23.7. Compare the use of thin slices of a solid figure for computing areas and volumes, as illustrated by Archimedes' *Method*, Bhaskara's computation of the area of a sphere, and Zu Chongzhi's computation of the volume of a sphere. What differences among the three do you notice?
- 23.8. The algebra developed by the Muslims and Europeans focused on expressing the solution of an equation as an *algebraic expression* involving the coefficients. The Chinese method, as we have seen, emphasizes finding *numerical approximations* to the roots. What are the advantages and disadvantages of each approach?
- 23.9. Since the geometric problems of finding the size of a town from measurements taken in a ludicrously indirect manner cannot have been the motive for studying cubic equations, what was the actual motive? Why was the geometry introduced at all?

Traditional Japanese Mathematics

Japan adopted the Chinese system of writing, and along with it a huge amount of vocabulary. The establishment of Buddhism in Japan in the sixth century increased the rate of cultural importation from China and even from India. The courses of university instruction in mathematics in Japan were based on reading (in Chinese) the classics discussed in Chapter 22. In relation to Japan, the Koreans played a role as transmitters, passing on Chinese learning and inventions. This transmission process began in 553–554 when two Korean scholars, Wang Lian-tung and Wang Puson, journeyed to Japan. For many centuries both the Koreans and the Japanese worked within the system of Chinese mathematics. The earliest records of new and original work in these countries date from the seventeenth century. By that time, mathematical activity was exploding in Europe, and Europeans had begun their long voyages of exploration and colonization. There was only a comparatively brief window of time during which indigenous mathematics independent of Western influence could grow up in Japan.

In the following synopsis, Japanese names are given with the family name first. A word of explanation is needed about the names, however. Most Chinese symbols (*kanji* in Japanese) have at least two readings in Japanese. For example, the symbol read as *CHŪ* (*middle*) in the Japanese word for China (*CHŪGOKU*, the “Middle Kingdom”) is also read as *naka* in the surname Tanaka (“Middlefield”). These variant readings often cause trouble in names from the past, so that one cannot always be sure how a name was pronounced. As Mikami (p. viii) says, “We read Seki Kōwa, although his personal name Kōwa should have been read Takakazu.” One symbol in that name is now read as *KŌ*, but apparently was once also read as *taka*. These are the so-called *ON* (Chinese) and *kun* (Japanese) readings of the same symbol. The *kun* reading of this symbol does not seem to exist any longer; it means, incidentally, *filial piety*.¹ A list of the names of some prominent mathematicians and their *kanji* rendering can be found at the end of the article by Martzloff (1990).

24.1. CHINESE INFLUENCE AND CALCULATING DEVICES

All the Japanese records now extant date from the time after Japan had adopted the Chinese writing system. Japanese mathematicians were for a time content to read the

¹When Japanese words are rendered in the Latin alphabet, it is customary to capitalize the *ON* pronunciations and write the *kun* in lowercase. The words *wasan* and *sangaku* that we shall be introducing are both *ON* readings, but we shall omit the capitalization.

Chinese classics. In 701 the emperor Monbu established a university system in which the mathematical part of the curriculum consisted of the Chinese Ten Classics. Some of these are no longer known, but the *Zhou Bi Suan Jing*, *Sun Zi Suan Jing*, *Jiu Zhang Suan Shu*, and *Hai Dao Suan Jing*, discussed in the two preceding chapters, were among them. Japan was disunited for many centuries after this early encounter with Chinese culture, and the mathematics that later grew up was the result of a reintroduction in the sixteenth and seventeenth centuries. Evidence of Chinese influence is seen in the mechanical methods of calculation used for centuries—counting rods, counting boards, and the abacus, which played an especially important role in Japan.

The abacus (*suan pan*) was invented in China, probably in the fourteenth century, when methods of computing with counting rods had become so efficient that the rods themselves were a hindrance to the performance of the computation. From China the invention passed to Korea, where it was known as the *sanbob*. Because it did not prove useful in Korean business, it did not become widespread there. It passed on to Japan, where it is known as the *soroban*. The Japanese made two technical improvements in the abacus: (1) They replaced the round beads by beads with sharp edges, which are easier to manipulate; and (2) they eliminated the superfluous second 5-bead on each string.

24.2. JAPANESE MATHEMATICIANS AND THEIR WORKS

A nineteenth-century Japanese historian reported that in the late sixteenth century, the ruling lord Hideyoshi sent the scholar Mōri Shigeyoshi (dates unknown, also known as Mōri Kambei) to China to learn mathematics. According to the story, the Chinese ignored the emissary because he was not of noble birth. When he returned to Japan and reported this fact, Hideyoshi conferred noble status on him and sent him back. Unfortunately, his second visit to China coincided with Hideyoshi's unsuccessful attempt to invade Korea, which made his emissary unwelcome in China. Mōri Shigeyoshi did not return to Japan until after the death of Hideyoshi. When he did return (in the early seventeenth century), he brought the abacus with him. Whether this story is true or not, it is a fact that Mōri Shigeyoshi was one of the most influential early Japanese mathematicians. He wrote several treatises, all of which have been lost, but his work led to a great flowering of mathematical activity in seventeenth-century Japan through the work of his students. This mathematics was known as *wasan*, and written using Chinese characters. The word *wasan* is written with two Chinese characters. The first is *WA*, a character used to denote Japanese-style work in arts, crafts, and cuisine; it means literally *harmony*. The second is *SAN*, meaning calculation, the same Chinese symbol that represents *suan* in the many Chinese classics mentioned above.² Murata (1994, p. 105) notes that the primary concern in *wasan* was to obtain elegant results, even when those results required very complicated calculations, and that “many *Wasanists* were men of fine arts rather than men of mathematics in the European sense.”

One unusual feature of mathematics during the Tokugawa Era from 1600 to 1868 was the choice of outlets in which to publish. Rather than writing letters to other scholars, or publishing in journals, as was common in Europe at this period, Japanese mathematicians would write books with problems in them to challenge other mathematicians. A phenomenon unique to Japanese mathematics is the tradition of *sangaku* (computational framed pictures),

²The modern Japanese word for mathematics is *SŪGAKU*, meaning literally *number theory*.

which were hung at Buddhist and Shintō shrines as votive plaques containing geometric problems leading to equations of higher degree. We shall discuss one of these in more detail below.

According to Murata, the stimulus for the development of *wasan* came largely from two Chinese classics: the *Suan Fa Tong Zong* by Cheng Dawei, published in 1592, and the older treatise *Suan Shu Chimeng (Introduction to Mathematical Studies)* by Zhu Shijie (ca. 1260–ca. 1320), published in 1299. The latter was particularly important, since it came with no explanatory notes and a rebellion in China had made communication with Chinese scholars difficult. By the time this treatise was understood, the Japanese mathematicians had progressed beyond its contents.

24.2.1. Yoshida Koyu

Mōri Shigeyoshi trained three outstanding students during his lifetime, of whom we shall discuss only the first. This student was Yoshida Koyu (Yoshida Mitsuyoshi, 1598–1672). Being handicapped in his studies at first by his weakness in Chinese, Yoshida Koyu devoted extra effort to this language in order to read the *Suan Fa Tong Zong*. Having read this book, Yoshida Koyu made rapid progress in mathematics and soon excelled even Mōri Shigeyoshi himself. Eventually, he was called to the court of a nobleman as a tutor in mathematics. In 1627 Yoshida Koyu wrote a textbook in Japanese, the *Jinkō-ki (Treatise on Large and Small Numbers)*, based on the *Suan Fa Tong Zong*. This work helped to popularize the abacus in Japan. It concluded with a list of challenge questions and thereby stimulated further work. These problems were solved in a later treatise, which, in turn, posed new mathematical problems to be solved; this was the beginning of a tradition of posing and solving problems that lasted for 150 years.

24.2.2. Seki Kōwa and Takebe Kenkō

One figure in seventeenth-century Japanese mathematics stands far above all others, a genius who is frequently compared with Archimedes, Newton, and Gauss.³ His name was Seki Kōwa (Takakazu)—the *wa* is the same symbol found in *wasan*—and he was born around 1640, making him a contemporary of Newton and Leibniz. The stories told of him resemble stories told about other mathematical geniuses. For example, one of his biographers says that at the age of five, Seki Kōwa pointed out errors in a computation that was being discussed by his elders. A similar story is told about Gauss. Being the child of a samurai father and adopted by a noble family, Seki Kōwa had access to books. He was mostly self-educated in mathematics, having paid little attention to those who tried to instruct him; in this respect he resembles Newton. Like Newton, he served as an advisor on high finance to the government, becoming examiner of accounts to the lord of Kosu. Unlike Newton, however, he was a popular teacher and physically vigorous. He became a shogunate samurai and master of ceremonies in the household of the Shogun. He died at the age of 66, leaving no direct heirs. His tomb in the Buddhist cemetery in Tokyo was rebuilt 80 years after his death

³His biography suggests that the real comparison should be with Pythagoras, since he assembled a devoted following, and his followers were inclined to attribute results to him even when his direct influence could not be established. Newton and Gauss were not “people persons,” and Gauss hated teaching. But Seki Kōwa had a close relationship with his students.

by mathematicians of his school. His pedagogical activity earned him the title of *Sansei*, meaning *Arithmetical Sage*, a title that was carved on his tomb. Although he published very little during his lifetime, his work became known through his teaching activity, and he is said to have left copious notebooks.

Seki Kōwa made profound contributions to several areas of mathematics, in some cases anticipating results that were being obtained independently in Europe about this time. According to Mikami (p. 160), he kept his technique a secret from the world at large, but apparently he confided it to his pupil Takebe Kenkō (Takebe Katahiro, 1664–1739). Some scholars say that Takebe Kenkō refused to divulge the secret, saying, “I fear that one whose knowledge is so limited as mine would tend to misrepresent its significance.” However, other scholars claim that Takebe Kenkō did write an exposition of the latter method and that it amounts to the principles of cancelation and transposition. These two scholars, together with Takebe Kenkō’s brother, compiled a 20-volume encyclopedia, the *Taisei Sankyō* (*Great Mathematical Treatise*), containing all the mathematics known in their day.

Takebe Kenkō also wrote a book that is unique in its time and place, bearing the title *Tetsujutsu Sankyō* (roughly, *The Art of Doing Mathematics*, published in 1722), in which he speculated on the metaphysics of mathematical concepts and the kind of psychology needed to solve different types of mathematical problems (Murata, 1994, pp. 107–108).

In Japan, knowledge of the achievements of Western mathematicians became widespread in the late nineteenth century, while the flow of knowledge in the opposite direction has taken longer. A book entitled *The Theory of Determinants in the Historical Order of Development*, which is a catalog of papers on the subject with commentaries, was published by the South African mathematician Thomas Muir (1844–1934) in 1905. Although this book consists of four volumes totaling some 2000 pages, it does not mention Seki Kōwa, the true discoverer of determinants!

24.2.3. The Modern Era in Japan

In the seventeenth century, the Tokugawa shoguns adopted a policy of exclusion vis-à-vis the West, one that could be enforced in an island kingdom such as Japan. Commercial contacts with the Dutch, however, resulted in some cultural penetration, and Western mathematical advances came to be known little by little in Japan. By the time Japan was opened to the West in the mid-nineteenth century, Japanese mathematicians were already aware of many European topics of investigation. In joining the community of nations for trade and politics, Japan also joined it intellectually. In the early nineteenth century, Japanese mathematicians were writing about such questions as the rectification of the ellipse, a subject of interest in Europe at the same period. By the end of the nineteenth century, there were several Japanese mathematical journals publishing (in European languages) mathematical work comparable to what was being done in Europe at the same period, and a few European scholars were already reading these journals to see what advances were being made by the Japanese. In the twentieth century the number of Japanese works being read in the West multiplied, and Japanese mathematicians such as Gorō Shimura (b. 1930), Shōshichi Kobayashi (b. 1932), and many others have been among the leaders in nearly every field of mathematics.

24.3. JAPANESE GEOMETRY AND ALGEBRA

During the seventeenth and eighteenth centuries there was a tradition of geometric challenge problems in Japan. The geometric problems usually involved combinations of simple figures

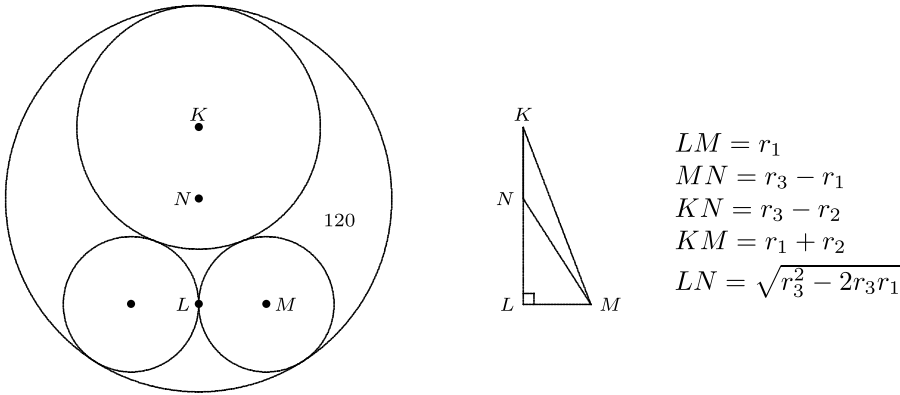


Figure 24.1. Sawaguchi Kazuyuki's first problem.

whose areas or volumes were known but which were arranged in such a way that finding their parts became an intricate problem in algebra. The word *algebra* needs to be emphasized here. The challenge in these problems was only superficially geometric; it was largely algebraic. The challenge was much greater to the Japanese mathematicians of the time than it is to us, since they did not have what we know as trigonometry. (They did have a rudimentary trigonometry, but they solved most problems using just the Pythagorean theorem.) We begin our discussion of this era by mentioning a few of the challenge problems. Afterward, we shall briefly discuss the infinitesimal methods used to solve the problems of measuring arcs, areas, and volumes in spheres.

One impetus to the development of mathematics in Japan came with the arrival of the Chinese “celestial element method” (*tian yuan shu*). This name was given to the unknown in an equation by Li Ye (1192–1279, also known as Li Zhi) in his 1248 treatise *Ceyuan Haijing* (*Sea Mirror of Circle Measurements*, see Mikami, 1913, p. 81).⁴ This term passed to Korea as *ch'onwonsul* and thence to Japan as *tengen jutsu*, which also means “celestial element method.” This Chinese algebra became part of the standard Japanese curriculum before the seventeenth century.

Fifteen problems were published by Sawaguchi Kazuyuki (dates unknown) in his 1670 work *Kokon Sampō-ki* (*Ancient and Modern Mathematics*). As an example of the complexity of these problems, consider the first of them. In this problem there are three circles, each externally tangent to the other two and internally tangent to a fourth circle, as in Fig. 24.1. The diameters of two of the enclosed circles are equal and the third enclosed circle has a diameter five units larger. The area inside the enclosing circle and outside the three smaller circles is 120 square units. The problem is to compute the diameters of all four circles. Without a computer algebra system, most people, even nowadays, would not wish to attempt to solve this problem. As Fig. 24.1 shows, the problem leads to the simultaneous equations

$$\begin{aligned} r_1 + \frac{5}{2} &= r_2, \\ 2\pi r_1^2 + \pi r_2^2 + 120 &= \pi r_3^2, \\ 4r_2^2 r_3 + 2r_1 r_2 r_3 + r_1 r_2^3 + r_1 r_3^2 &= 4r_2 r_3^2, \end{aligned}$$

⁴According to Libbrecht (1973, pp. 345–346), the same word had been used in a rather different and obscure sense by Qin Jiushao a year earlier in his *Shu Shu Jiu Zhang*.

where r_1 , r_2 , and r_3 are the radii of the circles. The last of these relations results from applying the Pythagorean theorem first to the triangle LMN to get LM and then to KLM .

This problem was solved by Seki Kōwa (Smith and Mikami, 1914, pp. 96–97). In case Seki Kōwa's prowess in setting up and solving equations was not clear from his solution of this problem, we note that he also solved the fourteenth of these problems, the "quadrilateral problem" (see below), which allegedly led to an equation of degree 1458. Although the procedure was a mechanical one, using counting boards, prodigious concentration must have been required to execute it. What a chess player Seki Kōwa could have been! As Mikami (1913, p. 160) remarks, "Perseverance and hard study were a part of the spirit that characterized Japanese mathematics of the old times."

24.3.1. Determinants

Seki Kōwa is given the credit for inventing one of the central ideas of modern mathematics: determinants. He introduced this subject in 1683 in *Kai Fukudai no Hō (Method of Solving Fukudai Problems)*.⁵ Nowadays, determinants are usually introduced in connection with linear equations, but Seki Kōwa developed them in relation to equations of higher degree as well. The method is explained as follows.

Suppose that we are trying to solve two simultaneous quadratic equations

$$\begin{aligned}ax^2 + bx + c &= 0, \\a'x^2 + b'x + c' &= 0.\end{aligned}$$

When we eliminate x^2 , we find the linear equation

$$(a'b - ab')x + (a'c - ac') = 0.$$

Similarly, if we eliminate the constant term from the original equations and then divide by x , we find

$$(ac' - a'c)x + (bc' - b'c) = 0.$$

Thus from two quadratic equations we have derived two linear equations. Seki Kōwa called this process *tatamu (folding)*.

We have written out expressions for the simple 2×2 determinants here. For example,

$$\begin{vmatrix} a & c \\ a' & c' \end{vmatrix} = ac' - a'c;$$

but, as everyone knows, the full expanded expressions for determinants are cumbersome even for the 3×3 case. It is therefore important to know ways of simplifying such determinants, using the structural properties we now call the *multilinear property* and the *alternating property*. Seki Kōwa knew how to make use of the multilinear property to take out a common factor from a given row. He not only formulated the concept of a determinant

⁵The word *fukudai* seems to be related to *fukugen suru*, meaning *reconstruct* or *restore*. According to Smith and Mikami (1914, p. 124), Seki Kōwa's school offered five levels of diploma, the third of which was called the *fukudai menkyo (fukudai license)* because it involved knowledge of determinants.

but also knew many of their properties, including how to determine which terms are positive and which are negative in the expansion of a determinant. It is interesting that determinants were introduced in Europe around the same time (1693, by Leibniz), but in a comparatively limited context.⁶ As Smith and Mikami (1914, p. 125) say,

It is evident that Seki was not only the discoverer but that he had a much broader idea than that of his great German contemporary.

Determinants are not the only topic on which Seki was on a par with the European mathematicians of his time. In one of his works, the *Katsuyō Sampō* (*Compendium of the Major Computational Rules*) published posthumously in 1712 (the year before the publication of James Bernoulli's *Ars conjectandi*), Seki gave a table of what are now called *Bernoulli numbers*. (See Smith and Mikami, 1914, p. 108.)

24.3.2. The Challenge Problems

As mentioned above, in 1627 Yoshida Koyu (1598–1672) wrote the *Jinkō-ki*, concluding it with a list of challenge questions. Here are some of those questions:

1. *There is a log of precious wood 18 feet long whose bases are 5 feet and $2\frac{1}{2}$ feet in circumference. Into what lengths should it be cut to trisect the volume?*
2. *There have been excavated 560 measures of earth, which are to be used for the base of a building. The base is to be 3 measures square and the building is to be 9 measures high. Required, the size of the upper base.*
3. *There is a mound of earth in the shape of a frustum of a circular cone. The circumferences of the bases are 40 measures and 120 measures and the mound is 6 measures high. If 1200 measures of earth are taken evenly off the top, what will be the height?*
4. *A circular piece of land 100 [linear] measures in diameter is to be divided among three persons so that they shall receive 2900, 2500, and 2500 [square] measures, respectively. Required, the lengths of the chords and the altitudes of the segments.*

Seki Kōwa solved a geometric problem that would challenge even the best algebraist today. It was the fourteenth in a list of challenge problems posed by Sawaguchi Kazuyuki: *There is a quadrilateral whose sides and diagonals are u , v , w , x , y , and z [as shown in Fig. 24.2].*

It is given that

$$\begin{aligned} z^3 - u^3 &= 271, \\ u^3 - v^3 &= 217, \\ v^3 - y^3 &= 60.8, \\ y^3 - w^3 &= 326.2, \\ w^3 - x^3 &= 61. \end{aligned}$$

Required, to find the values of u , v , w , x , y , z .

⁶A recent paper whose author has not yet given permission for citation gives strong evidence that much of matrix theory as we now know it was common throughout Eurasia during the Medieval period, and that in fact the West may actually have learned about determinants from China.

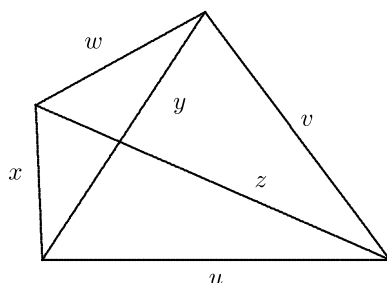


Figure 24.2. Sawaguchi Kazuyuki's quadrilateral problem.

The fact that the six quantities are the sides and diagonals of a quadrilateral provides one equation that they must satisfy, namely:

$$u^4 w^2 + x^2 (v^4 + w^2 y^2 - v^2 (w^2 - x^2 + y^2)) - (y^2 (w^2 + x^2 - y^2) + v^2 (-w^2 + x^2 + y^2)) z^2 + y^2 z^4 - u^2 (v^2 (w^2 + x^2 - y^2) + w^2 (-w^2 + x^2 + y^2) + (w^2 - x^2 + y^2) z^2) = 0.$$

This equation, together with the five given conditions, provides a complete set of equations for the six quantities. However, Seki Kōwa's explanation, which is only a sketch, does not mention this sixth equation, so it may be that what he solved was the indeterminate problem given by the other five equations. That, however, would be rather strange, since then the quadrilateral would play no role whatsoever in the problem. Whatever the case, it is known that such equations were solved numerically by the Chinese using a counting board.

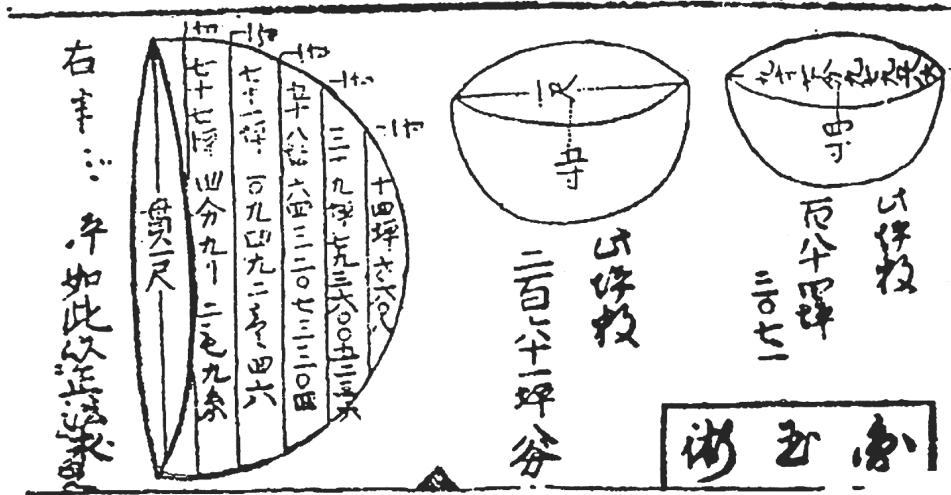
The quadrilateral problem of Sawaguchi Kazuyuki led to an equation of degree 1458, solved by Seki Kōwa (who was Sawaguchi Kazuyuki's teacher). Again we emphasize that this problem—like many of the problems in the *sangaku* plaques and like many problems studied by mathematicians from Mesopotamia to India—seems to be inspired by the desire to do some complicated algebra rather than by any pressing geometric need.

24.3.3. Beginnings of the Calculus in Japan

By the end of the seventeenth century the *wasanists* were beginning to use techniques that resemble the infinitesimal methods being used in Europe about this time. Of course, in one sense Zu Chongzhi had used some principles of calculus 1000 years earlier in his application of Cavalieri's principle to find the volume of a sphere. The intuitive basis of the principle is that equals added to equals yield equal sums, and a solid can be thought of as the sum of its horizontal sections. It isn't really, of course. No finite sum of areas and no limit of such a sum can ever have positive volume. Students in calculus courses learn to compute volumes using approximating sums that are very thin prisms, but not infinitely thin.

In Japan these techniques were first applied in the area called *yenri* (circle theory),⁷ a topic that had been studied extensively in China. The idea of approximating by shells or disks,

⁷The symbol for circle (*yen*) is also the symbol for the Japanese unit of currency; in Japan, it is actually pronounced "en."



Isomura Kittoku’s computation of the volume of a sphere. Copyright © Stock Montage.

now a basic part of courses in calculus, can be seen in the 1684 edition of the *Ketsugi-shō* (*Combination Book*), first published in 1660 by Isomura Kittoku (dates unknown).

Isomura Kittoku explained the method as follows (Mikami, 1913, p. 204):

If we cut a sphere of diameter 1 foot into 10,000 slices, the thickness of each slice is 0.0001 feet, which will be something like that of a very thin paper. Finding in this way the volume of each of them, we sum up the results, 10,000 in number, when we get 523.6 measures [that is, a volume of 0.5236 cubic feet]. Besides, it is true, there are small incommensurable parts, which are neglected.

Since the volume of this sphere is $\pi/6 \approx 0.5235987756$, this technique is quite accurate. All that is required to get it is the formula for the area of a circle, plus the Pythagorean theorem to determine the square of the radius of each slice. Except for the numerical value of π , all this can be done in integer arithmetic, with no error, provided the radius is an integer. The technique of obtaining extraordinary precision and using it to perform numerical experiments that provide the basis for inductive reasoning is very close to the technique used by Bhaskara II (see Chapter 21) to compute the surface area of a sphere. It also appears in a remarkable infinite series attributed to Takebe Kenkō, as we shall now see.

Takebe Kenkō’s method of rectifying the circle was based on a relation, which he apparently discovered in 1722, between the square of half of an arc, the height h of the arc,⁸ and the diameter d of the circle. Here is his own description of this discovery, as explained by Smith and Mikami (pp. 1914, 147–149). He began with height $h = 0.000001 = 10^{-6}$ and $d = 10$, finding the square of the arc geometrically with accuracy to 53 decimal places. The value of the square of this arc is

$$0.00001\ 00000\ 00333\ 33335\ 11111\ 12253\ 96833\ 52381\ 01394\ 90188\ 203 + .$$

⁸This height is called the *sagitta* (arrow) by lens grinders, a name first bestowed on it in India. It is now called the *versed sine* in mathematics. In our terms, it is $1 - \cos \theta$ times the radius.

According to Smith and Mikami (1914, p. 148), the value given by Takebe Kenkō was

0.00000 00000 33333 35111 11225 39690 66667 28234 77694 79595 875 + .

But this value does not fit with the procedure followed by Takebe Kenkō; it does not even yield the correct first approximation. The figure given by Smith and Mikami appears to represent the value obtained by Takebe Kenkō *after* the first approximation was subtracted, but with the result multiplied by the square of the diameter.⁹ In appreciating Takebe Kenkō's method, the first problem to be solved is the source of this extremely accurate measurement of the circle. According to Smith and Mikami (1914, p. 148), Takebe Kenkō said that the computation was given in two other works, both of which are now lost.

The first clue that strikes us in this connection is the seemingly strange choice of the *square* of the arc rather than the arc itself. Why would it be easier to compute the square of the arc than the arc itself? An answer readily comes to mind: Half of the arc is approximated by its chord, and the chord is one side of a convenient right triangle. In fact, the chord of half of an arc is the mean proportional between the diameter of the circle and the height of the full arc, so that in this case it is $\sqrt{dh} = \sqrt{10^{-5}}$. When we square it, we get just $dh = 10^{-5}$, which acts as Takebe Kenkō's first approximation. That result suggests that the length of the arc was reached by repeatedly bisecting the arc, taking the chord as an approximation. This hypothesis gains plausibility, since it is known that this technique had been used earlier to approximate π . Since $a^2 = 4(a/2)^2$, it was only necessary to find the square of half the arc, then multiply by 4. The ratio of the chord to the diameter is even easier to handle, especially since Takebe Kenkō has taken the diameter to be 10. If x is the square of this ratio for a given chord, the square of ratio for the chord of half of the arc is $(1 - \sqrt{1-x})/2$. In other words, the iterative process $x \mapsto (1 - \sqrt{1-x})/2$ makes the bisection easy. If we were dealing with the arc instead of its square, each step in that process would involve two square roots instead of one. Even as it is, Takebe Kenkō must have been a calculating genius to iterate this process enough times to get 53 decimal places of accuracy without making any errors. The result of 50 applications yields a ratio which, multiplied by $100 \cdot 4^{50}$, is

0.00001 00000 00333 33335 11111 12253 96833 52381 01131 94822 94294 362 + .

This number of iterations gives 38 decimal places of accuracy. Even with this plausible method of procedure, it still strains credibility that Takebe Kenkō achieved the claimed precision. However, let us pass on to the rest of his method.

After the first approximation hd is subtracted, the new error is 10^{-12} times $0.3333333 \dots$, which suggests that the next correction should be $10^{-12}/3$. But this is exactly $h^2/3$, in other words $h/(3d)$ times the first term. When it is subtracted from the previously corrected value, the new error is

$10^{-19} \cdot 0.17777 77892 06350 01904 76806 15685 4870 + .$

⁹Even so, there is one 3 missing at the beginning and, after it is restored, the accuracy is "only" 33 decimal places. That precision, however, would have been all that Takebe Kenkō needed to compute the four corrections he claimed to have computed.

The long string of 7's here suggests that this number is 10^{-19} times $\frac{1}{10} + \frac{7}{90} = \frac{16}{90} = \frac{8}{45}$, which is $(8h)/(15d)$ times the previous correction. By continuing for a few more terms, Takebe Kenkō was able to observe a pattern: The corrections are obtained by multiplying successively by $h/(3d)$, $(8h)/(15d)$, $(9h)/(14d)$, $(32h)/(45d)$, $(25h)/(33d)$, Some sensitivity to the factorization of integers is necessary to see the recursive operation: multiplication by $(h/d)[2n^2/(n+1)(2n+1)]$. Putting these corrections together as an infinite series leads to the expression

$$\frac{a^2}{4} = dh \left[1 + \sum_{n=1}^{\infty} \frac{2^{2n+1}(n!)^2}{(2n+2)!} \cdot \left(\frac{h}{d}\right)^n \right]$$

when the full arc has length a .

In using this numerical approach, Takebe Kenkō had reached his conclusion inductively. This induction was based on a faith (which turns out to be justified) that the successive approximations are rational numbers that satisfy a fairly simple recursive formula. As you probably know, the power series for the sine, cosine, exponential, and logarithm have this happy property, but the series for the tangent, for example, does not.

This series solves the problem of rectification of the circle and hence all problems that depend on knowing the value of π . In modern terms the series given by Takebe Kenkō represents the function

$$\left(d \arcsin \left(\sqrt{\frac{h}{d}} \right) \right)^2.$$

Takebe Kenkō's discovery of this result in 1722 falls between the discovery of the power series for the arcsine function by Newton in 1676 and its publication by Euler in 1737. For an arc of 60° , we have $d = 2r$ and $h = r(1 - \sqrt{3}/2)$; and with these values, ten terms of the series will approximate $\pi^2/36$ to 15 decimal places (when $r = 1$).

24.4. SANGAKU

The shoguns of the Tokugawa Era (1600–1868) concentrated their foreign policy on relations with China and held Western visitors at arms length, with the result that Japan was nearly closed to the Western world for 250 years. During this time a form of mathematics known as *sangaku* (literally, *framed computations*) arose. As mentioned above, the *sangaku* were votive tablets containing mathematical problems posted at Buddhist and Shintō shrines as a challenge to others and an expression of piety. A comprehensive exhibition of *sangaku* from many parts of Japan was organized in 2005 at the Nagoya Museum of Science, and a book with full color illustrations from that exhibit was published. (See Fukagawa, 2005.)

In 1806, Ehara Masanori (dates unknown) hung a colorful picture of the diagram sketched in Fig. 24.3 at the Atsuta Shintō shrine in Owari Province¹⁰ It was subsequently lost, but not before the Owari scholar Kitagawa Mōko (1763–1839) made a pilgrimage to the shrine

¹⁰Since 1868, Owari has formed the western part of Ai Chi Prefecture. It includes the city of Nagoya, where the Atsuta shrine is located.

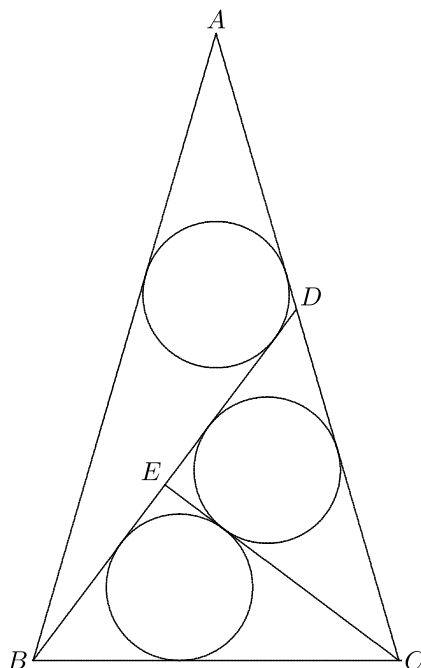


Figure 24.3. A *sangaku* problem.

and solved the problem. From his description of the problem and his solution, the people at the shrine were able to reconstruct the replica of the original document shown in Fig. 24.4.

Here is the problem as reconstructed: In Fig. 24.3, the triangle ABC is isosceles. The line BD has been drawn meeting the bisector CE of $\angle C$ at right angles. Obviously then, $CD = BC$, and the triangles CDE and CBE are congruent. Therefore their inscribed circles are also congruent. The remarkable thing about this particular triangle is that the circle inscribed in triangle BDA is also congruent to the other two circles. The problem asks for the common radius of the three circles (in terms of the line CE).

This problem is discussed in the book by Fukagawa and Rothman (2008). The problem is stated on pages 194–196, and the solution by Kitagawa Mōko is given on pages 212–216.

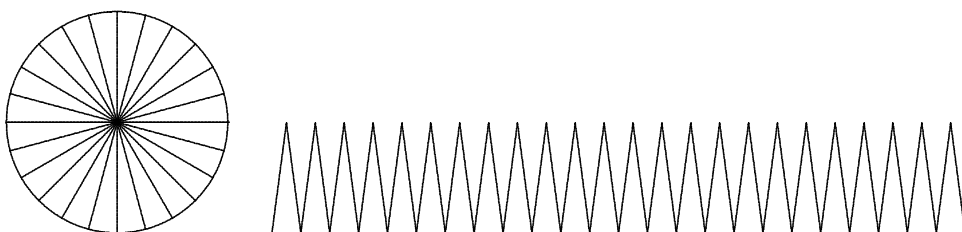


Figure 24.4. A disk cut into sectors and opened up.

24.4.1. Analysis

It should be observed first of all that there is precisely one base angle for the isosceles triangle such that these three circles are all the same size, and it is somewhere between 60° and 90° . The third circle degenerates to a point when the base angles are $\frac{\pi}{3}$ (the triangle is equilateral), while the radius of the other two circles is $\frac{\overline{BC}}{\sqrt{3} + \sqrt{7+4\sqrt{3}}}$. On the other hand, when the base angles become $\frac{\pi}{2}$ (the triangle degenerates to an infinite strip), the radius of the upper circle is $\frac{\overline{BC}}{2}$ while that of the two lower triangles is $\frac{1}{2}\overline{BC}(\sqrt{2} - 1)$. Thus, the upper circle grows to be bigger than the two lower circles as the base angles increase. Hence, there is precisely *one* angle for which this condition can be met (and it is between 73° and 74°).

PROBLEMS AND QUESTIONS

Mathematical Problems

- 24.1.** Given the very broad hint that $z = 10$, $u = 9$, $v = 8$, $w = 5$, and $x = 4$, solve the quadrilateral problem to get an exact expression for y , and exact expressions for which the data of the problem (the numbers 60.8 and 326.2) are approximations.
- 24.2.** Early on, Japanese mathematicians believed the area of the sphere to be one-fourth the square of the circumference, that is, $\pi^2 r^2$ rather than the true value $4\pi r^2$. This value was stated in the first (1660) edition of Isomura Kittoku's *Ketsugi-Sho* and corrected in a later edition. Smith and Mikami (1914, p. 75) suggest a way in which this belief might have appeared plausible. To explain it, we first need to see an example in which the same line of reasoning really does work.

By imagining a circle sliced like a pie into a very large number of very thin pieces, one can imagine it cut open and all the pieces laid out next to one another, as shown in Fig. 24.4. Because these pieces are very thin, their bases are such small arcs of the circle that each base resembles a straight line. Neglecting a very tiny error, we can say that if there are n pieces, the base of each piece is a straight line of length $2\pi r/n$. The sectors are then essentially triangles of height r (because of their thinness) and hence area $(1/2) \cdot (2\pi r^2)/n$. Since there are n of them, the total area is πr^2 . This heuristic argument (exactly what Archimedes stated in the letter accompanying his paper on the surface of the sphere) gives the correct result. In fact, this very figure appears in a Japanese work from 1698 (Smith and Mikami, 1914 p. 131).

Now imagine a hemispherical bowl covering the pie. If the slices are extended upward so as to slice the bowl into equally thin segments, and those sectors are then straightened out and arranged like the sectors of the pie, they also will have bases equal to $\frac{2\pi r}{n}$, but their height will be one-fourth of the circumference, in other words, $\pi r/2$, giving a total area for the hemisphere of $(1/2) \cdot \pi^2 r^2$. Since the area is $2\pi r^2$, this would imply that $\pi = 4$. How much error would there be in taking $\pi = 4$?

- 24.3.** Problem 41 of Isomura Kittoku (Smith and Mikami, 1914, p. 68) is given as the following: *There is a log 18 feet long, whose ends have diameters 1 foot and 2.6 feet. It is wound spirally with 75 feet of string, the coils being 2.5 feet apart. How many coils are there?*

This problem seems to be overdetermined, since the length of the log and the distance between the coils would already determine the number of coils. The maximum number would be 7.2, but the total length of them would be only a little over 40 feet. Omitting the assumption that the coils are 2.5 feet apart, solve the problem that remains.

Historical Questions

- 24.4. When and how did mathematical research arise in Japan, and how did mathematicians learn their subject?
- 24.5. What areas of mathematics became specialties in Japan, and what innovations arose there?
- 24.6. What is *sangaku*?

Questions for Reflection

- 24.7. What is wrong with the reasoning above that leads to the conclusion that the area of a sphere is one-fourth the square of the circumference?
- 24.8. What is the justification for the statement by the historian of mathematics T. Murata that Japanese mathematics (*wasan*) was not a science but an art?
- 24.9. Why might Seki Kōwa and other Japanese mathematicians have wanted to keep their methods secret, and why did their students, such as Takebe Kenkō, honor that wish?

ISLAMIC MATHEMATICS, 800–1500

The next three chapters constitute a sampling of mathematical advances from a civilization that flourished over a huge region from Spain to India during the period known in Europe as the Middle Ages.

Contents of Part V

1. Chapter 25 (Overview of Islamic Mathematics) establishes the cultural and historical context of the subject and introduces the major figures and their works to be discussed in the two chapters that follows.
2. Chapter 26 (Islamic Number Theory and Algebra) discusses number theory and algebra from al-Khwarizmi and Thabit ibn-Qurra through Omar Khayyam.
3. Chapter 27 (Islamic Geometry) is devoted to Islamic advances in geometry.

Overview of Islamic Mathematics

As mentioned in the introduction to this part of the present history, the most important advances in science and mathematics in the West from 700 to 1300 CE came in the lands under Muslim rule.

25.1. A BRIEF SKETCH OF THE ISLAMIC CIVILIZATION

Starting as a small and persecuted sect in the early seventh century, by mid-century the Muslims had expanded by conquest as far as Persia. They then turned West and conquered Egypt, all of the Mediterranean coast of Africa, and the island of Sicily.

25.1.1. The Umayyads

A palace revolution among the Islamic leaders led to the triumph of the first dynasty, the Umayyad (sometimes spelled Omniad) in the year 660. Under the Umayyads, Muslim expansion continued around the Mediterranean coast and eastward as far as India. This expansion was checked by the Byzantine Empire at the Battle of Constantinople in 717. In the West a Muslim general named Tarik led an army into Spain, giving his name to the mountain at the southern tip of Spain—Jabal Tarik, known in English as Gibraltar. The Muslim expansion in the West was halted by the Franks under Charles Martel at the Battle of Tours in 732. In 750 another revolution resulted in the overthrow of the Umayyad Dynasty and its replacement in the East by the Abbasid Dynasty. The Umayyads remained in power in Spain, however, a region known during this time as the Caliphate of Cordoba.

From the early ninth century on, scholars working under the rule of the caliphs formed a unique tradition within the story of mathematics, sharing a common literature of mathematical classics, communicating with one another, and working to extend the achievements of their predecessors. Their achievements were considerable, and Europeans from the eleventh century on were eager to learn about them and apply them. Because the origin of Islam lies in the Arabic-speaking world, and its holy text is written in Arabic, most of the documents produced within this tradition were written in Arabic by scholars for whom the Arabic language was either native or learned at school. Some non-Arabic writers, especially in the early years, adopted Arabic names. As with Mesopotamian and Greek mathematics,

there is some inaccuracy in any name one might choose to refer to this tradition. Should it be called *Arabic mathematics* because of the language most commonly used to write it, or *Islamic mathematics* because Islam is the most obvious feature that most of the writers had in common? Whichever name the reader prefers, the important thing is to grasp what the name signifies—that is, the specific sets of questions and problems the mathematicians studied and the approaches they had for solving them.

25.1.2. The Abbasids

Al-Mansur, the second of the Abbasid caliphs, built the capital of the new dynasty, the city of Baghdad, on the Tigris River. Both the Abbasids and the Umayyads cultivated science and the arts, and mathematics made advances in both the Eastern and Western parts of the Islamic world. The story of Islamic mathematics begins in the city of Baghdad in the reign of two caliphs. The first of these was Harun al-Raschid (786–809), a contemporary of Charlemagne. The second is the son of Harun al-Raschid, al-Mamun (813–833), whose court life provided the setting of the *Thousand and One Nights*.

25.1.3. The Turkish and Mongol Conquests

Near the end of the tenth century a group of Turkish nomads called Seljuks migrated from Asia into the Abbasid territory and converted to Islam. Gradually the Seljuks began to seize territory from the Abbasids, and in 1055 they occupied Baghdad. It was their advance into Palestine that provoked the First Crusade in 1096. The Crusades, which established a Christian-ruled enclave in Palestine, were another source of continuing disruption throughout the twelfth century and even later. The Seljuks left the Abbasids as the nominal rulers of the empire, but in the thirteenth century both Abbasids and Seljuks were conquered by the same Mongols who had earlier overrun Russia and China. The Mongol conquest of Iraq was particularly devastating, since it resulted in the destruction of the irrigation system that had supported the economy of the area for thousands of years. As in China, the Mongol rule was short-lived and was succeeded by another conquest, this time by the Ottoman Turks, who also conquered Constantinople in 1453 and remained a threat to Europe until the nineteenth century. While it lasted, the vast Mongol Empire transmitted mathematical works and ideas over prodigious distances. In particular, astronomical treatises came into China from Persia, along with Arabic numerals (Li and Du, 1987, pp. 171–174).

25.1.4. The Islamic Influence on Science

The portion of the Islamic empire around the Mediterranean Sea was secure from invasion for three hundred years in the East and six hundred in Spain. During this period, Islamic mathematicians assimilated the science and mathematics of their predecessors and made their own unique additions and modifications to what they inherited. For many centuries they read the works of Archimedes, Apollonius, and Euclid and advanced beyond the work of these illustrious Greek mathematicians. The Greek mathematicians, however, were not the only influence on them. From earliest times the Caliph was in diplomatic contact with India, and one of Harun Al-Raschid's contributions was to obtain translations from Sanskrit into Arabic of the works of Aryabhata, Brahmagupta, and others. Some of the translators took

the occasion to write their own mathematical works, and so began the Islamic contribution to mathematics.¹

In addition to the Arabic translations that preserved many Greek works of which the originals have been lost, the modern world has inherited a considerable amount of scientific and mathematical literature in Arabic. This language has given us many words relating to science, such as *alcohol*, *alchemy*, *almanac*, *zenith*, and the mysterious names of the stars such as *Altair*, *Aldebaran*, *Algol*, and *Betelgeuse*. In Spain, the libraries were incomparably richer than those in northern Europe until well past the year 1000, and many scholars from the Christian countries of Europe came there to translate Arabic works into Latin.²

Thus, from the end of the eighth century through the period referred to as Medieval in European history, the Umayyad and Abbasid Caliphates, centered in what is now Spain and Iraq respectively, produced an artistically and scientifically advanced culture, with works on mathematics, physics, chemistry, and medicine written in Arabic, the common language of scholars throughout the Muslim world. Persian, Hebrew, and other languages were also used by scholars working in this predominantly Muslim culture. The label *Islamic mathematics* that we are going to use has one important disadvantage, since we certainly have no wish to imply that mathematical results valid in one religion are not valid in another. Yet the alternative, *Arabic mathematics*, also does not seem to fit as well as the corresponding label *Greek mathematics*, in which the majority of the major authors had Greek as their native language.

25.2. ISLAMIC SCIENCE IN GENERAL

The religion of Islam calls for prayers facing Mecca at specified times of the day. That alone would be sufficient motive for studying astronomy and geography. Since the Muslim calendar is lunar rather than lunisolar, religious feasts and fasts are easy to keep track of. Since Islam forbids representation of the human form in paintings, mosques are always decorated with abstract geometric patterns (see Özdural, 2000). The study of this *ornamental geometry* has interesting connections with the theory of transformation groups. Unfortunately, we do not have space to pursue this interesting topic, nor the equally fascinating subject of the astrolabe, which was highly developed as an almanac and surveying tool by Muslim scholars.

25.2.1. Hindu and Hellenistic Influences

According to Colebrooke (1817, pp. lxiv–lxv), in the year 773 CE, al-Mansur, the second caliph of the Abbasid Dynasty, who ruled from 754 to 775, received at his court a Hindu scholar bearing a book on astronomy referred to in Arabic as *Sind-hind* (most likely, *Siddhanta*). Al-Mansur had this book translated into Arabic. No copies survive. It was once conjectured that this book was the *Brahmasphutasiddhanta* mentioned in Chapter 21, but

¹Plofker, (2009, p. 258), however, cautions against assuming that algebra among the Muslims had its roots in these Sanskrit works, pointing out that the works written in Arabic do not use negative numbers.

²Constantinople, which had preserved its independence, continued a mathematical tradition until the fifteenth century, and it also was an important source of ancient works for the Europeans. Unfortunately, we do not have space to discuss the details of that recovery effort.

Plofker (2009, p. 256) cites two papers of Pingree (1968, 1970) in rejecting this conjecture. This book was used for some decades, and an abridgement was made in the early ninth century, during the reign of al-Mamun (caliph from 813 to 833), by Muhammad ibn Musa al-Khwarizmi (ca. 780–850), who also wrote his own treatise on astronomy based on the Hindu work and the work of Ptolemy. Al-Mamun founded a “House of Wisdom” (*Bait al-Hikma*) in Baghdad, the capital of his empire. This institution was much like the Library at Alexandria, a place of scholarship, analogous to a modern research institute.

In the early days of this scientific culture, one of the concerns of the scholars was to find and translate into Arabic as many scientific works as possible. The effort made by Islamic rulers, administrators, and merchants to acquire and translate Hindu and Hellenistic texts was prodigious. The works had first to be located, a job requiring much travel and expense. Next, they needed to be understood and adequately translated; that work required a great deal of labor and time, often involving many people. The world is much indebted to the scholars who undertook this work, for two reasons. First, some of the original works have been lost, and only their Arabic translations survive.³ Second, the translators, inspired by the work they were translating, wrote original works of their own. The mechanism of this two-part process has been described by Berggren (1990, p. 35):

Muslim scientists and patrons were the main actors in the acquisition of Hellenistic science inasmuch as it was they who initiated the process, who bore the costs, whose scholarly interests dictated the choice of material to be translated and on whom fell the burden of finding an intellectual home for the newly acquired material within the Islamic *dār al-‘ilm* (“abode of learning”).

The acquisitions were extensive, and we have space for only a partial enumeration of them. Some of the major ones were listed by Berggren (2002). They include Euclid’s *Elements*, *Data*, and *Phænomena*, Ptolemy’s *Syntaxis* (which became the *Almagest* as a result) and his *Geography*, many of Archimedes’ works and commentaries on them, and Apollonius’ *Conics*.

The development process as it affected the *Conics* of Apollonius was described by Berggren (1990, pp. 27–28). This work was used to analyze the astrolabe in the ninth century and to trisect the angle and construct a regular heptagon in the tenth century. It continued to be used down through the thirteenth century in the theory of optics, for solving cubic equations and to study the rainbow. To the two categories that we have called acquisition and development, Berggren adds the process of editing the texts to systematize them, and he emphasizes the very important role of mathematical philosophy or criticism engaged in by Muslim mathematicians. They speculated on and debated Euclid’s parallel postulate, for example, thereby continuing a discussion that began among the ancient Greeks and continued for 2000 years until it was finally settled in the nineteenth century.

The scale of the Muslim scientific schools is amazing when looked at in comparison with the populations and the general level of economic development of the time. Here is an excerpt from a letter of the Persian mathematician al-Kashi (d. 1429) to his father, describing

³Toomer (1984b) points out that in the case of Ptolemy’s *Optics* the Arabic translation has also been lost, and only a Latin translation from the Arabic survives. As Toomer notes, some of the most interesting works were not available in Spain and Sicily, where medieval scholars went to translate Arabic and Hebrew manuscripts into Latin.

the life of Samarkand, in Uzbekistan, where the great astronomer Ulugh Beg (1374–1449), grandson of the conqueror Timur the Lame, had established his observatory (Bagheri, 1997, p. 243):

His Royal Majesty had donated a charitable gift. . . amounting to thirty thousand. . . dinars, of which ten thousand had been ordered to be given to students. [The names of the recipients] were written down; [thus] ten thousand-odd students steadily engaged in learning and teaching, and qualifying for a financial aid, were listed. . . Among them there are five hundred persons who have begun [to study] mathematics. His Royal Majesty the World-Conqueror, may God perpetuate his reign, has been engaged in this art. . . for the last twelve years.

25.3. SOME MUSLIM MATHEMATICIANS AND THEIR WORKS

We now survey some of the more important mathematicians who lived and worked under the rule of the caliphs.

25.3.1. Muhammad ibn Musa al-Khwarizmi

This scholar, who lived from approximately 790 to 850, translated a number of Greek works into Arabic but is best remembered for his *Hisab al-Jabr w'al-Mugabalah (Book of Completion and Reduction)*. The word *completion* (or *restoration*) here (*al-jabr*) is the source of the modern word *algebra*. It refers to the operation of keeping an equation in balance by adding or subtracting the same terms on both sides of an equation, as in the process of completing the square. The word *reduction* refers to the cancelation of a common factor from the two sides of an equation. The author came to be called simply al-Khwarizmi, which may be the name of his home town (although this is not certain); this name gave us another important term in modern mathematics, *algorithm*.

The integration of intellectual interests with religious piety that we saw in the case of the Hindus is a trait also possessed by the Muslims. Al-Khwarizmi introduces his algebra book with a hymn of praise of Allah and then dedicates his book to al-Mamun:

That fondness for science, by which God has distinguished the Imam al-Mamun, the Commander of the Faithful. . . , that affability and condescension which he shows to the learned, that promptitude with which he protects and supports them in the elucidation of obscurities and in the removal of difficulties—has encouraged me to compose a short work on Calculating by (the rules of) Completion and Reduction, confining it to what is easiest and most useful in arithmetic, such as men constantly require in cases of inheritance, legacies, partition, law-suits, and trade, and in all their dealings with one another, or where the measuring of lands, the digging of canals, geometrical computation, and other objects of various sorts. . . My confidence rests with God, in this as in every thing, and in Him I put my trust. . . May His blessing descend upon all the prophets and heavenly messengers. [Rosen, 1831, pp. 3–4]

25.3.2. Thabit ibn-Qurra

The Sabian (star-worshipping) sect centered in the town of Harran in what is now Turkey produced an outstanding mathematician/astronomer in the person of Thabit ibn-Qurra (826–901). Being trilingual (besides his native Syriac, he spoke Arabic and Greek), he was invited to Baghdad to study mathematics. His mathematical and linguistic skills procured him work

translating Greek treatises into Arabic, including Euclid's *Elements*. He was a pioneer in the application of arithmetic operations to ratios of geometric quantities, which is the essence of the idea of a real number. The same idea occurred to René Descartes (1596–1650) and was published in his famous work on analytic geometry. It is likely that Descartes drew some inspiration from the works of the fourteenth-century Bishop of Lisieux Nicole d'Oresme (1323–1382); Oresme, in turn, is likely to have read translations from the Arabic. Hence it is possible that our modern concept of a real number owes something to the genius of Thabit ibn-Qurra. He also wrote on mechanics, geometry, and number theory.

25.3.3. Abu Kamil

Although nothing is known of the life of Abu Kamil (ca. 850–930), he is the author of certain books on algebra, geometry, and number theory that influenced both Islamic and European mathematics. Many of his problems were reproduced in the work of Leonardo of Pisa (Fibonacci, 1170–1250).

25.3.4. Al-Battani

Another Sabian from Harran, Abu Abdallah Muhammad al-Battani, known in Latin translation as Albategnius, seems to have abandoned the Sabian beliefs of his parents and converted to Islam. That, at least, is what has been inferred from his Muslim name. Since he himself reported making astronomical observations in the year 877, he must have been born some time during the 850s. He died around 929. He worked in al-Raqqah in what is now Syria, on the Euphrates River. His best-known work is the *Kitab al-Zij (Book of Astronomy)*. The word *zij* apparently comes from Persian, where it means a certain strand in a rug.

The first three of the 57 chapters of al-Battani's book contain a development of trigonometry using sines, one that has been claimed to be independent of the work of Aryabhata I. Obviously, however, he must have known something about Aryabhata's works, or else he would have invented an Arabic name for the sine, instead of borrowing the Sanskrit *j-y-b* that will be discussed in Chapter 27.

25.3.5. Abu'l Wafa

Muhammad Abu'l Wafa (940–998) was born in Khorasan (now in Iran) and died in Baghdad. He was an astronomer–mathematician who translated Greek works and commented on them. In addition he wrote a number of works on practical arithmetic and geometry. According to Rashid (1994), his book of practical arithmetic for scribes and merchants begins with the claim that it “comprises all that an experienced or novice, subordinate or chief in arithmetic needs to know” in relation to taxes, business transactions, civil administration, measurements, and “all other practices. . . which are useful to them in their daily life.”

25.3.6. Ibn al-Haytham

Abu Ali al-Hasan ibn al-Haytham (965–1039), known in the West as Alhazen, was a natural philosopher who worked in the tradition of Aristotle. He continued the speculation on the parallel postulate, offering a proof of it that was, of course, flawed. He is famous for *Alhazen's problem* in optics, which is to determine the point on a reflecting spherical surface at which a light ray from one given point P will be reflected to a second given point Q .

25.3.7. Al-Biruni

Abu Arrayhan al-Biruni (973–1048), was an astronomer, geographer, and mathematician who as a young man worked out the mathematics of maps of the earth. Civil wars in the area where he lived (Uzbekistan and Afghanistan) made him into a wanderer, and he came into contact with astronomers in Persia and Iraq. He was a prolific writer. According to the *Dictionary of Scientific Biography*, he wrote what would now be well over 10,000 pages of texts during his lifetime, on geography, geometry, arithmetic, and astronomy.

25.3.8. Omar Khayyam

The Persian mathematician Omar Khayyam, also known as Umar al-Khayyam, was born in 1044 and died in 1123. He is thought to be the same person who wrote the famous skeptical and hedonistic poem known as the *Rubaiyat (Quatrains)*, but not all scholars agree that the two are the same. Since he lived in the turbulent time of the invasion of the Seljuk Turks, his life was not easy, and he could not devote himself wholeheartedly to scholarship. Even so, he advanced algebra beyond the linear and quadratic equations discussed in al-Khwarizmi's book and speculated on the foundations of geometry. He explained his motivation for doing mathematics in the preface to his *Algebra*. Like the Japanese *wasanists*, he was inspired by questions left open by his predecessors. As with al-Khwarizmi, this intellectual curiosity is linked with piety and with gratitude to the patron who supported his work:

In the name of God, gracious and merciful! Praise be to God, Lord of all Worlds, a happy end to those who are pious, and ill-will to none but the merciless. May blessings repose upon the prophets, especially upon Mohammed and all his holy descendants.

One of the branches of knowledge needed in that division of philosophy known as mathematics is the science of completion and reduction, which aims at the determination of numerical and geometrical unknowns. Parts of this science deal with certain very difficult introductory theorems, the solution of which has eluded most of those who have attempted it. . . I have always been very anxious to investigate all types of theorems and to distinguish those that can be solved in each species, giving proofs for my distinctions, because I know how urgently this is needed in the solution of difficult problems. However, I have not been able to find time to complete this work, or to concentrate my thoughts on it, hindered as I have been by troublesome obstacles. [Kasir, 1931, pp. 43–44]

25.3.9. Sharaf al-Tusi

Sharaf al-Din al-Tusi (ca. 1135–1213) is best remembered for work on cubic equations. Judging from the name al-Tusi, he must have been born near the town of Tus in northeastern Iran. Like Omar Khayyam, he lived in turbulent times. The Seljuk Turks had captured Damascus in 1154 and established their capital in that city. Sharaf al-Tusi is known to have taught there around 1165 and to have moved from there to Aleppo (also in Syria).

25.3.10. Nasir al-Tusi

Nasir al-Din al-Tusi (1201–1274) had the misfortune to live during the time of the westward expansion of the Mongols, who subdued Russia during the 1240s and then went on to conquer Baghdad in 1258. Al-Tusi himself joined the Mongols and was able to continue

his scholarly work under the new ruler Hulegu, grandson of Genghis Khan. Hulegu, who died in 1265, conquered and ruled Iraq and Persia over the last decade of his life, taking the title *Ilkhan* when he declared himself ruler of Persia. A generation later the Ilkhan rulers converted from Buddhism to Islam. Hulegu built al-Tusi an observatory at Maragheh, a city in the Azerbaijan region of Persia that Hulegu had made his seat of government. Here al-Tusi was able to improve on the earlier astronomical theory of Ptolemy, in connection with which he developed both plane and spherical trigonometry into much more sophisticated subjects than they had been previously, including the statement that the sides of triangles are proportional to the sines of the angles opposite them. Because of his influence, the loss of Baghdad was less of a blow to Islamic science than it would otherwise have been. Nevertheless, the constant invasions had the effect of greatly reducing the vitality and the quantity of research. Al-Tusi played an important role in the flow of mathematical ideas back into India after the Muslim invasion of that country; it was his revised and commented edition of Euclid's *Elements* that was mainly studied (De Young, 1995, p. 144).

QUESTIONS

Historical Questions

- 25.1. Describe the general history of Muslim expansion and political decline over the period from the eighth to fifteenth centuries.
- 25.2. Who were the major mathematicians working within the world of Islamic scholarship during this time, and what topics did they develop?
- 25.3. What justifications do al-Khwarizmi and Omar Khayyam give in the prefaces to their work for the algebra that they develop?
- 25.4. In what way was Nasir al-Tusi's trigonometry an advance on the subject as inherited from the Hindu mathematicians?

Questions for Reflection

- 25.5. How did the conquests by different groups of Muslims affect the course of scholarship in the conquered areas (Spain, Mesopotamia, India, China)?
- 25.6. How did the Islamic injunction against representation of the human body in art influence art and architecture in the Islamic countries?
- 25.7. If one needs to pray facing Mecca while living in (say) Chicago, how is "facing Mecca" to be interpreted? How can one work out how to face Mecca from Chicago? This problem is not difficult to solve using spherical trigonometry. To find out how al-Biruni solved it, see the book by Berggren (1986, pp. 182–186).
- 25.8. An expository short book (Brett, Feldman, and Sentlowitz, 1974) giving some history of mathematics contains the following statement (p. 41) about Islamic mathematics:

It is often said that the Arabs were learned but not original; thus, they played the role of preservation rather than invention of knowledge. Even if we believe this description of them,

we must be forever grateful for the benevolent custody by the Moslems of the world's intellectual possessions which might otherwise have been lost forever in the mire of the Dark Ages.

It is to their credit that the authors do not endorse what they report as a popular impression of the Islamic world. Most Western historians have given that culture credit for outstanding achievements in art, literature, and science. The charge of a lack of creativity is also sometimes made against Byzantine Empire contemporaneous with the Islamic—again unfairly, since its wealth and geographical range were tiny by comparison with the world of Islam. Here, for example, is what the British philosopher Bertrand Russell (1872–1969) said about it (1945, p. xvi):

In the Eastern Empire, Greek civilization, in a desiccated form, survived, as in a museum, till the fall of Constantinople in 1453, but nothing of importance to the world came out of Constantinople except an artistic tradition and Justinian's Codes of Roman law.

Russell did not disparage Islamic science, but in his own area of philosophy, he did tend to look down on Islamic scholarship, saying (p. 417)

Arabic philosophy is not important as original thought. Men like Avicenna [ibn Sina (980–1037), Persian physician] and Averroes [ibn Rushd (1126–1198), Spanish philosopher], are essentially commentators.

Whether this negative opinion is justified or not is a matter for philosophers to discuss, and any opinion by a non-philosopher would be rash. Let it be said, however, that *important* is a word whose meaning may vary from one philosopher to another.

Western writers, it is true, sometimes overlook Islamic contributions and slight them with silence. For example, Kline (1953, p. 93) in discussing the Medieval period in Europe, says:

The progress that was made during this period was contributed by the Hindus and Arabs. . .

He goes on to list a number of Hindu mathematical discoveries, and then finishes with this comment:

These and other Hindu contributions were acquired by the Arabs who transmitted them to Europeans.

Kline was simply writing carelessly here. He knew better, and he gave more detailed discussions of the Islamic contributions in his later, encyclopedic work (Kline, 1972).

Giving these authors the benefit of the doubt, since no one can discuss every single meritorious deed in any history, how is it possible to write concisely, yet with fairness to the subject? If you were editing the works just quoted, how would you advise the authors to recast these sentences?

Islamic Number Theory and Algebra

It is well known that the numerals used all over the world today are an inheritance from both the Hindu and Arabic mathematicians of 1000 years ago. The Hindu idea of using nine symbols in a place-value system was known in what is now Iraq in the late seventh century, before that area became part of the Muslim Empire. In the late eighth century a scholar from India came to the court of Caliph al-Mansur with a work on Hindu astronomy using these numerals, and this work was translated into Arabic. An Arabic treatise on these numbers, containing the first known discussion of decimal fractions, was written by al-Uqlidisi (ca. 920–ca. 980).

Having inherited works from the time of Mesopotamia and also Greek and Hindu works that used the sexagesimal system in astronomy, the Muslim mathematicians of a thousand years ago also used that system. The sexagesimal system did not yield immediately to its decimal rival, and the technique of place-value computation developed in parallel in the two systems. Ifrah (2000, pp. 539–555) gives a detailed description of the long resistance to the new system. The sexagesimal system is mentioned in Arabic works of Abu'l-Wafa and Kushar ben Laban (ca. 971–1029). It continued to appear in Arabic texts through the time of al-Kashi (1427), although the decimal system also occurs in the work of al-Kashi.¹

Some implementations of the decimal system require crossing out or erasing in the process of computation, and that was considered a disadvantage. Nevertheless, the superiority of decimal notation in computation was recognized early. For example, al-Daffa (1973, pp. 56–57) mentions that there are manuscripts still extant dating to the twelfth century, in which multiplication is performed by the very efficient method illustrated in Fig. 26.1 for the multiplication $524 \cdot 783 = 410,292$.

26.1. NUMBER THEORY

The Muslims continued the work of Diophantus in number theory. Abu Kamil wrote a book on “indeterminate problems” in which he studied quadratic Diophantine equations and systems of such equations in two variables. The first 38 problems that he studied are

¹In addition to the sexagesimal and decimal systems, the Muslim mathematicians used an elaborate system of finger reckoning.

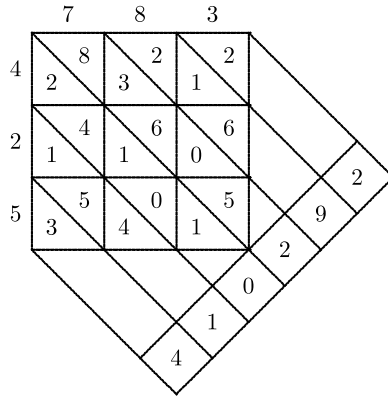


Figure 26.1. The computation $524 \times 783 = 410,292$.

arranged in order of coefficients, exponents, and signs, making a systematic exposition of these equations. Later scholars noted the astonishing fact that the first 25 of these equations are what are now known as algebraic curves of genus 0, while the last 13 are of genus 1, even though the concept of genus of an algebraic curve is a nineteenth-century invention (Baigozhina, 1995).

Muslim mathematicians also went beyond what is in Euclid and Nicomachus, generalizing perfect numbers. In a series of articles, Rashed [see, for example, Rashed (1989)] pointed out that a large amount of theory of abundant, deficient, and perfect numbers was assembled in the ninth century by Thabit ibn-Qurra and others and that ibn al-Haytham (965–1040) was the first to state and attempt to prove that Euclid’s formula gives all the even perfect numbers. Thabit ibn Qurra made an interesting contribution to the theory of amicable numbers. A pair of numbers is said to be *amicable* if each is the sum of the parts (proper divisors) of the other. The smallest such pair of numbers is 220 and 284. Although these numbers are not discussed by Euclid or Nicomachus, the commentator Iamblichus (see Dickson, 1919, p. 38) ascribed this notion to Pythagoras, who is reported as saying, “A friend is another self.” This definition of a friend is given by Aristotle in his *Nicomachean Ethics* (Bekker, 2000).

In Chapter 9, we discussed the only known way of generating perfect numbers, namely the Euclidean formula $2^{n-1}(2^n - 1)$, whenever $2^n - 1$ is a prime. Thabit ibn-Qurra found a similar way of generating pairs of amicable numbers. His formula is

$$2^n(3 \cdot 2^n - 1)(3 \cdot 2^{n-1} - 1) \text{ and } 2^n(9 \cdot 2^{2n-1} - 1),$$

whenever $3 \cdot 2^n - 1$, $3 \cdot 2^{n-1} - 1$, and $9 \cdot 2^{2n-1} - 1$ are all prime. The case $n = 2$ gives the pair 220 and 284. Whatever one may think about the impracticality of amicable numbers, there is no denying that Thabit’s discovery indicates very profound insight into the divisibility properties of numbers. It is very difficult to imagine how he could have discovered this result. A conjecture, which cannot be summarized in a few lines, can be found in the article by Brentjes and Hogendijk (1989).

It is not clear how many new cases can be generated from this formula, but there definitely are some. For example, when $n = 4$, we obtain the amicable pair $17, 296 = 16 \cdot 23 \cdot 47$ and $18, 416 = 16 \cdot 1151$. Hogendijk (1985) gives Thabit ibn-Qurra’s proof of his criterion for

amicable numbers and points out that the case $n = 7$ generates the pair 9,363,584 and 9,437,056, which first appeared in Arabic texts of the fourteenth century.

Unlike some other number-theory problems such as the Chinese remainder theorem, which arose in a genuinely practical context, the theory of amicable numbers is an offshoot of the theory of perfect numbers, which was already a completely “useless” topic from the beginning. It did not seem useless to the people who developed it, however. According to M. Cantor (1880, p. 631), the tenth-century mystic al-Majriti recommended as a love potion writing the numbers on two sheets of paper and eating the number 284, while causing the beloved to eat the number 220. He claimed to have verified the effectiveness of this charm by personal experience! Dickson (1919, p. 39) mentions the Jewish scholar Abraham Azulai (1570–1643), who described a work purportedly by the ninth-century commentator Rau Nachshon, in which the gift of 220 sheep and 220 goats that Jacob sent to his brother Esau as a peace offering (Genesis 32:14) is connected with the concept of amicable numbers.² In any case, although their theory seems more complicated, amicable numbers are easier to find than perfect numbers. Euler alone found 62 pairs of them (see Erdős and Dudley, 1983).

Another advance on the Greeks can be found in the work of Kamal al-Din al-Farisi, a Persian mathematician who died around 1320. According to Aġargün and Fletcher (1994), he wrote the treatise *Memorandum for Friends Explaining the Proof of Amicability*, whose purpose was to give a new proof of Thabit ibn-Qurra’s theorem. Proposition 1 in this work asserts the existence (but not uniqueness) of a prime decomposition for every number. Propositions 4 and 5 assert that this decomposition is unique, that two distinct products of primes cannot be equal.

26.2. ALGEBRA

It has always been recognized that Europe received algebra from the Muslims. As we have already said, the word *algebra* (*al-jabr*) is an Arabic word meaning *completion* or *restoration*.³ Its origins in the Muslim world date from the ninth century, in the work of al-Khwarizmi, as is well established.⁴

What is less certain is how much of al-Khwarizmi’s algebra was original with him and how much he learned from Hindu sources. According to Colebrooke (1817, pp. lxiv–lxxx), he was well versed in Sanskrit and translated a treatise on Hindu computation⁵ into Arabic

²The peace offering was necessary because Jacob had tricked Esau out of his inheritance. But if the gift was symbolic and associated with amicable numbers, this interpretation seems to imply that Esau was obligated to give Jacob 284 sheep and 284 goats. Perhaps there was an ulterior motive in the gift!

³Gandz (1926) presented a different theory of the origin of the term *algebra*, according to which the word is not even of Arabic origin, despite its Arabic appearance. To the extent that the majority rules in such matters, this alternative theory is heavily outvoted by the one just described.

⁴Colebrooke (1817, p. lxxiii) noted that a manuscript of this work dated 1342 was in the Bodleian Library at Oxford. Obviously, this manuscript could not be checked out, and Colebrooke complained that the library’s restrictions “preclude the study of any book which it contains, by a person not enured to the temperature of apartments unvisited by artificial warmth.” If he worked in the library in 1816, his complaint would be understandable: Due to volcanic ash in the atmosphere, there was no summer that year. This manuscript is the source that Rosen (1831) translated and reproduced.

⁵It is apparently this work that brought al-Khwarizmi’s name into European languages in the form *algorism*, now *algorithm*. A Latin manuscript of this work in the Cambridge University Library, dating to the thirteenth century, has been translated into English (Crossley and Henry, 1990).

at the request of Caliph al-Mamun. Colebrooke cites the Italian writer Pietro Cossali,⁶ who presented the alternatives that al-Khwarizmi learned algebra either from the Greeks or the Hindus and opted for the Hindus. These alternatives are a false dichotomy. We need not conclude that al-Khwarizmi took everything from the Hindus or that he invented everything himself. It is very likely that he expounded some material that he read in Sanskrit and added his own ideas to it. Rosen (1831, p. x) explains the difference in the preface to his edition of al-Khwarizmi's algebra text, saying that "at least the method which he follows in expounding his rules, as well as in showing their application, differs considerably from that of the Hindu mathematical writers."

Colebrooke also notes (p. lxxi) that Abu'l-Wafa wrote a translation or commentary on the *Arithmetica* of Diophantus. This work, however, is now lost. Apart from these possible influences of Greek and Hindu algebra, whose effect is difficult to measure, it appears that the progress of algebra in the Islamic world was an indigenous growth. We shall trace that growth through several of its most prominent representatives, starting with the man recognized as its originator, Muhammad ibn Musa al-Khwarizmi.

26.2.1. Al-Khwarizmi

Besides the words *algebra* and *algorithm*, there is a common English word whose use is traceable to Arabic influence (although it is not an Arabic word), namely *root* in the sense of a square or cube root or a root of an equation. The Greek picture of the square root was the side of a square, and the word *side* (*pleurá*) was used accordingly. The Muslim mathematicians apparently thought of the root as the part from which the equation was generated and used the word *jadhr* accordingly. According to al-Daffa (1977, p. 80), translations into Latin from Greek use the word *latus* while those from Arabic use *radix*. In English the word *side* lost out completely in the competition.

Al-Khwarizmi's numbers correspond to what we call positive real numbers. Theoretically, such a number could be defined by any convergent sequence of rational numbers, but in practice some rule is needed to generate the terms of the sequence. For that reason, it is more accurate to describe al-Khwarizmi's numbers as positive *algebraic numbers*, since all of his numbers are generated by equations with rational coefficients. The absence of negative numbers prevented al-Khwarizmi from writing all quadratic equations in the single form "squares plus roots plus numbers equal zero" ($ax^2 + bx + c = 0$). Instead, he had to consider three basic cases and two others, in which either the square or linear term is missing. He described the solution of "squares plus roots equal numbers" by the example of "a square plus 10 roots equal 39 *dirhems*." (A *dirhem* is a unit of money.) Al-Khwarizmi's solution of this problem is to draw a square of unspecified size (the side of the square is the desired unknown) to represent the square (Fig. 26.2). To add 10 roots, he then attaches to each side a rectangle of length equal to the side of the square and width $2\frac{1}{2}$ (since $4 \cdot 2\frac{1}{2} = 10$). The resulting cross-shaped figure has, by the condition of the problem, area equal to 39. He then fills in the four corners of the figure (literally "completing the square"). The total area of these four squares is $4 \cdot (2\frac{1}{2})^2 = 25$. Since $39 + 25 = 64$, the completed square has side 8.

⁶Cossali's dates are 1748–1813. He was Bishop of Parma and author of *Origine, trasporto in Italia, primi progressi in essa dell' algebra* (*The Origins of Algebra and Its Transmission to Italy and Early Advancement There*), published in Parma in 1797.

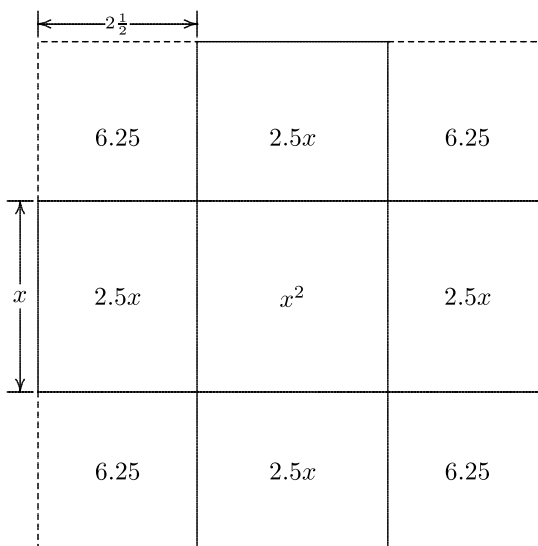


Figure 26.2. Al-Khwarizmi's solution of "square plus 10 roots equals 39 *dirhems*."

Since this square was obtained by adding rectangles of side $2\frac{1}{2}$ to each side of the original square, it follows that the original square had side 3.

This case is the one al-Khwarizmi considers first and is the simplest to understand. His figures for the other two cases of quadratic equations are more complicated, but all are based on a geometric illustration of the identity $((a + b)/2)^2 - ((a - b)/2)^2 = ab$.

Al-Khwarizmi did not consider any cubic equations. Roughly the first third of the book is devoted to various examples of pure mathematical problems leading to quadratic equations, causing the reader to be somewhat skeptical of his claim to be presenting the material needed in commerce and law. There are no genuine applications of quadratic equations in the book. Although quadratic equations have no practical applications (outside of technology, of course), there are occasions when a practical problem requires solving linear equations. Al-Khwarizmi found many such cases in problems of inheritance, which occupy more than half of his *Algebra*. Here is a sample:

A man dies, leaving two sons behind him, and bequeathing one-fifth of his property and one *dirhem* to a friend. He leaves 10 *dirhems* in property and one of the sons owes him 10 *dirhems*. How much does each legatee receive?

Although mathematics is cross-cultural, its applications are specific to the culture in which they are used. The difference between the modern solution of this legal problem and al-Khwarizmi's solution is considerable. Under modern law the man's estate would be considered to consist of 20 *dirhems*, the 10 *dirhems* cash on hand, and the 10 *dirhems* owed by one of the sons. The friend would be entitled to 5 *dirhems* (one-fifth plus one *dirhem*), and the indebted son would owe the estate 10 *dirhems*. His share of the estate would be one-half of the 15 *dirhems* left after the friend's share is taken out, or $7\frac{1}{2}$ *dirhems*. He would therefore have to pay $2\frac{1}{2}$ *dirhems* to the estate, providing it with cash on hand equal to $12\frac{1}{2}$ *dirhems*. His brother would receive $7\frac{1}{2}$ *dirhems*.

Now the notion of an estate as a legal entity that can owe and be owed money is a modern European one, alien to the world of al-Khwarizmi. Apparently, in al-Khwarizmi's time, money could be owed only to a *living person*. What principles are to be used for settling accounts in this case? Judging from the solution given by al-Khwarizmi, the estate is to consist of the 10 *dirhems* cash on hand, plus a *certain portion* (not all) of the debt the second son owed to his deceased father. This "certain portion" is the unknown in a linear equation and is the reason for invoking algebra in the solution. It is to be chosen so that *when the estate is distributed, the indebted son neither receives any more money nor owes any to the other heirs*. This condition leads to a linear equation. Al-Khwarizmi explains the solution as follows (we put the legal principle that provides the equation in capital letters):

Call the amount taken out of the debt *thing*. Add this to the property; the sum is 10 *dirhems* plus *thing*. Subtract one-fifth of this, since he has bequeathed one-fifth of his property to the friend. The remainder is 8 *dirhems* plus $\frac{4}{5}$ of *thing*. Then subtract the 1 *dirhem* extra that is bequeathed to the friend. There remain 7 *dirhems* and $\frac{4}{5}$ of *thing*. Divide this between the two sons. The portion of each of them is $3\frac{1}{2}$ *dirhems* plus $\frac{2}{5}$ of *thing*. THIS MUST BE EQUAL TO THING. Reduce it by subtracting $\frac{2}{5}$ of *thing* from *thing*. Then you have $\frac{3}{5}$ of *thing* equal to $3\frac{1}{2}$ *dirhems*. Form a complete *thing* by adding to this quantity $\frac{2}{3}$ of itself. Now $\frac{2}{3}$ of $3\frac{1}{2}$ *dirhems* is $2\frac{1}{3}$ *dirhems*, so that *thing* is $5\frac{5}{6}$ *dirhems*.

Rosen (1831, p. 133) suggested that the many arbitrary principles used in these problems were introduced by lawyers to protect the interests of next-of-kin against those of other legatees.

26.2.2. Abu Kamil

A commentary on al-Khwarizmi's *Algebra* was written by Abu Kamil.⁷ His exposition of the subject contained none of the legacy problems found in al-Khwarizmi's treatise, but after giving the basic rules of algebra, it listed 69 problems to be solved. For example, a paraphrase of Problem 10 is as follows:

The number 50 is divided by a certain number. If the divisor is increased by 3, the quotient decreases by $3\frac{3}{4}$. What is the divisor?

Abu Kamil is also noteworthy because many of his problems were copied by Leonardo of Pisa, one of the first to introduce the mathematics of the Muslims into Europe.

26.2.3. Omar Khayyam

Although al-Khwarizmi did not consider any equations of degree higher than 2, such equations were soon to be considered by Muslim mathematicians. A link between geometry and algebra appeared in the use of the rectangular hyperbola by Pappus to carry out the *neûsis* construction for trisecting an angle (see Section 3 of Chapter 11). Omar Khayyam (see Amir-Moez, 1963) realized that a large class of geometric problems of this type led to cubic

⁷A commentary on the commentary was written in Hebrew by the Italian Jewish scholar Mordecai Finzi (1440–1475). The present example is taken from the English translation of that work (Levey, 1966).

equations that could be solved using conic sections. His treatise on algebra⁸ was largely occupied with the classification and solution of cubic equations by this method. Before we discuss a general cubic equation solved by Omar Khayyam, we note one particular equation of this type that he posed and solved (Amir-Moez, 1963). That problem is to find the point on a circle such that the perpendicular from the point to a radius has the same ratio to the radius that the two segments into which it divides the radius have to each other.

If the radius is r and the length of the longer segment cut off on the radius is the unknown x , the equation to be satisfied is $x^3 + rx^2 + r^2x = r^3$. Without actually writing out this equation, Omar Khayyam showed that the geometric problem amounted to using the stated condition to find the second asymptote of a rectangular hyperbola, knowing one of its asymptotes and one point on the hyperbola. However, he regarded that analysis as merely an introduction to his real purpose, which was a discussion of the kinds of cubic equations that require conic sections for their solution. After a digression to classify these equations, he returned to the original problem and, finally, showed how to solve it using a rectangular hyperbola. He found the arc to be about 57° , so that $x \approx r \cos(57^\circ) = 0.544r$. Omar Khayyam described x as being about $30\frac{2}{3}$ pieces, that is, sixtieths of the radius.

Omar Khayyam did not have modern algebraic symbolism. Experience had evidently taught him that attempts to solve the general cubic equation by arithmetic and root extractions would not work in general. But he discovered that such an equation could be interpreted geometrically and solved by the use of conic sections. In applying those conic sections, he wrote in the language of Apollonius and Euclid, with the single exception of representing the lines as numbers. His classification of equations, like al-Khwarizmi's, is conditioned by the use of only positive numbers as data. For that reason his classification is even more complicated than al-Khwarizmi's, since he is considering cubic equations as well as quadratics. He lists 25 types of equations (Kasir, 1931, pp. 51–52), six of which do not involve any cubic terms.

By way of illustration, we shall consider the case of *cubes plus squares plus sides equal number*, or, as we would phrase it, $x^3 + ax^2 + bx = c$. In keeping with his geometric interpretation of magnitudes as line segments, Omar Khayyam had to regard the coefficient b as a square, so that we shall write b^2 rather than b . Similarly, he regarded the constant term as a solid, which without any loss of generality he considered to be a rectangular prism whose base was an area equal to the coefficient of the unknown. In keeping with this reduction we shall write b^2c instead of c . Thus Omar Khayyam was considering the equation $x^3 + ax^2 + b^2x = b^2c$, where a , b , and c are data for the problem, to be represented as lines. His solution is illustrated in Fig. 26.3. He drew a pair of perpendicular lines intersecting at a point O and marked off $OA = a$ and $OC = c$ in opposite directions on one of the lines and $OB = b$ on the other line. He then drew a semicircle having AC as diameter, the line DB through B perpendicular to OB (parallel to AC), and the rectangular hyperbola through C having DB and the extension of OB as asymptotes. This hyperbola intersects the semicircle in the point C and in a second point Z . From Z he drew ZP perpendicular to the extension of OB . This line ZP represented the solution of the cubic.

When it comes to actually producing a root by numerical procedures, Omar Khayyam's solution is circular, a mere restatement of the problem. He has broken the cubic equation into two quadratic equations in two unknowns, but any attempt to eliminate one of the two

⁸This treatise was little noticed in Europe until a French translation by Franz Woepcke (1827–1864) appeared in 1851 (Kasir, 1931, p. 7).

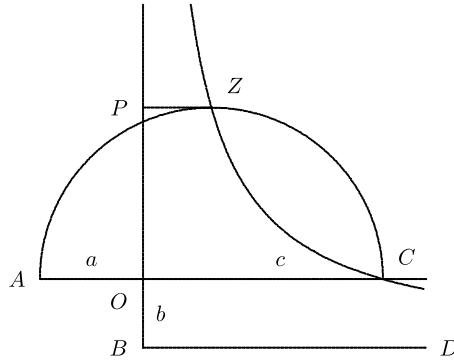


Figure 26.3. Omar Khayyam's solution of $x^3 + ax^2 + b^2x = b^2c$.

unknowns merely leads back to the original problem. In fact, no method of solution exists or can exist that reduces the solution of every cubic equation with real roots to the extraction of real square and cube roots of real numbers. What Omar Khayyam had created was an *analysis* of cubic equations using conic sections. He said that no matter how hard you look, you will never find a numerical solution “because whatever is obtained by conic sections cannot be obtained by arithmetic” (Amir-Moez, 1963, p. 336).

26.2.4. Sharaf al-Din al-Tusi

A generation after the death of Omar Khayyam, Sharaf al-Din al-Tusi wrote a treatise on equations in which he analyzed the cubic equation using methods that are surprisingly modern in appearance. This work has been discussed by Hogendijk (1989). Omar Khayyam had distinguished the eight types of cubic equations that always have a solution and five that could fail to have a solution. Al-Tusi provided a numerical method of solution for the first eight types that was essentially the Chinese method of solving cubic equations. He then turned to the five types that might have no (positive) solutions for some values of the data. As an example, one of these forms is

$$x^3 + ax^2 + c = bx.$$

For each of these cases, al-Tusi considered a particular value of x , which for this example is the value m satisfying

$$3m^2 + 2am = b.$$

Let us denote the positive root of this equation (the larger root, if there are two) by m . The reader will undoubtedly have noticed that the equation can be obtained by differentiating the original equation and setting x equal to m . The point m is thus in all cases a relative minimum of the difference of the left- and right-hand sides of the equation. That is precisely the property that al-Tusi wanted. Hogendijk comments that it is unlikely that al-Tusi had any concept of a derivative. In fact, the equation for m can be derived without calculus, by taking m as the value at which the minimum occurs, subtracting the values at x from the value at m , and dividing by $m - x$. The result is the inequality $m^2 + mx + x^2 + a(m + x) > b$ for $x > m$ and the opposite inequality for $x < m$. Therefore equality must hold when $x = m$,

that is, $3m^2 + 2am = b$, which is the condition given by al-Tusi.⁹ After finding the point m , al-Tusi concluded that there will be no solutions if the left-hand side of the equation is larger than the right-hand side when $x = m$. There will be one unique solution, namely $x = m$ if equality holds there. That left only the case in which the left-hand side was smaller than the right-hand side when $x = m$. For that case, he considered the auxiliary cubic equation

$$y^3 + py^2 = d,$$

where p and d were determined by the type of equation. The quantity d was the difference between the right- and left-hand sides of the equation at $x = m$, that is, $bm - m^3 - am^2 - c$ in the present case, with p equal to $3m + a$. Al-Tusi was replacing x with $y = x - m$ here. The procedure was precisely the method we know as Horner's method, and the linear term drops out because the condition by which m was chosen ordains that it be so. The equation in y was known to have a root because it was one of the other 13 types, which always have solutions. Thus, it followed that the original equation must also have a solution, $x = m + y$, where y was the root of the new equation. The added bonus was that a lower bound on m was obtained.

PROBLEMS AND QUESTIONS

Mathematical Problems

- 26.1.** Solve the following legacy problem from al-Khwarizmi's *Algebra*: *A woman dies and leaves her daughter, her mother, and her husband, and she bequeaths to some person as much as the share of her mother and to another as much as one-ninth of her entire capital. Find the share of each person.* It was understood from legal principles that the mother's share would be $\frac{2}{13}$ and the husband's $\frac{3}{13}$.
- 26.2.** Solve the problem of Abu Kamil in the text.
- 26.3.** Consider the cubic equation of Sharaf al-Tusi's third type, which we write as $x^3 + ax^2 - bx + c = 0$. Using the Chinese/Horner's method described in Chapter 22, show that if the first approximation is $x = m$, where m satisfies $3m^2 + 2am - b = 0$, then the equation to be satisfied at the second approximation is $y^3 + (3m + a)y^2 + (m^3 + am^2 - bm + c) = 0$. That is, carry out the algorithm for reduction and show that the process is

$$\begin{array}{r} 1 \\ a \\ -b \\ c \end{array} \longrightarrow \begin{array}{r} 1 \\ 3m + a \\ 3m^2 + 2am - b (= 0) \\ m^3 + am^2 - bm + c \end{array}$$

⁹This way of finding the minimum was also used by Fermat in the seventeenth century.

Historical Questions

- 26.4. What is the subject matter of al-Khwarizmi's *Algebra*, and what applications does it include?
- 26.5. How did Omar Khayyam solve cubic equations geometrically, and why does he adopt the geometric approach rather than a numerical one?
- 26.6. What refinements to the solution of the cubic equation are due to Sharaf al-Din al-Tusi?

Questions for Reflection

- 26.7. Why did al-Khwarizmi include a complete discussion of the solution of quadratic equations in his treatise when he had no applications for them at all?
- 26.8. Contrast the modern Western solution of the Islamic legacy problem discussed in the text with the solution of al-Khwarizmi. Is one solution "fairer" than the other? Can mathematics make any contribution to deciding what is fair?
- 26.9. Why did Omar Khayyam express the answer to a problem involving circles in *pieces* equal to one-sixtieth of the radius?

Islamic Geometry

In the Western world, most of the progress in geometry during the millennium that passed between the fall of the Western Roman Empire and the fall of the Eastern Empire occurred among the Muslim and Jewish mathematicians of Baghdad, Samarkand, Cordoba, and other places. This work had some features of Euclid's style and some of Heron's. Matvievskaya (1999) has studied the extensive commentaries on the tenth book of Euclid's *Elements* written by Muslim scholars from the ninth through twelfth centuries and concluded that while formally preserving a Euclidean distinction between magnitude and number, they actually operated with quadratic and quartic irrationals as if they were numbers.

27.1. THE PARALLEL POSTULATE

The Islamic mathematicians continued the later Hellenistic speculation on Euclid's parallel postulate. According to Sabra (1969), this topic came into Islamic mathematics through a commentary by Simplicius on Book 1 of the *Elements*, whose Greek original is lost, although an Arabic translation exists. In fact, Sabra found a manuscript that contains Simplicius' attempted proof. The reworking of this topic by Islamic mathematicians consisted of a criticism of Simplicius' argument followed by attempts to repair its defects. Gray (1989, pp. 42–54) presents a number of these arguments, beginning with the ninth-century mathematician al-Gauhari. Al-Gauhari attempted to show that two lines constructed so as to be parallel, as in Proposition 27 of Book 1 of the *Elements* must also be equidistant at all points. If he had succeeded, he would indeed have proved the parallel postulate.

27.2. THABIT IBN-QURRA

Thabit ibn-Qurra, whose revision of the Arabic translation of Euclid became a standard in the Muslim world, also joined the debate over the parallel postulate. According to Gray (1989, pp. 43–44), he considered a solid body moving without rotating so that one of its points P traverses a straight line. He claimed that the other points in the body would also move along straight lines, and obviously they would remain equidistant from the line generated by the point P . By regarding these lines as completed loci, he avoided a certain objection that could be made to a later argument of ibn al-Haytham, discussed below. Thabit

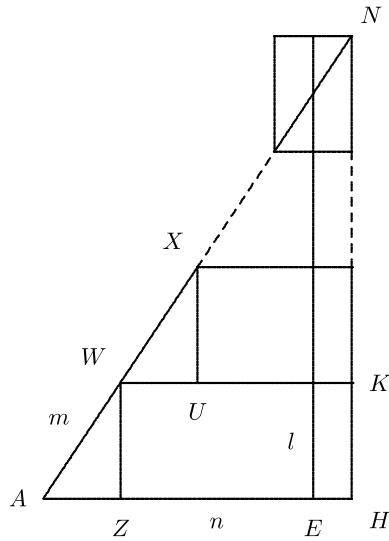


Figure 27.1. Thabit ibn-Qurra’s attempted proof of the parallel postulate.

ibn-Qurra’s work on this problem was ground-breaking in a number of ways, anticipating much that is usually credited to the eighteenth-century mathematicians Lambert and Saccheri. He proved, for example, that if a quadrilateral has two equal adjacent angles, and the sides not common to these two angles are equal, then the other two angles are also equal to each other. In the case when the equal angles are right angles, such a figure is called—unjustly, we may say—a *Saccheri quadrilateral*, after Giovanni Saccheri (1667–1733), who, like Thabit ibn-Qurra, developed it in an attempt to prove the parallel postulate. Gray prefers to call it a *Thabit quadrilateral*, and we shall use this name. Thabit ibn-Qurra’s proof amounted to the claim that a perpendicular drawn from one leg of such a quadrilateral to the opposite leg would also be perpendicular to the leg from which it was drawn. Such a figure, a quadrilateral having three right angles, or half of a Thabit quadrilateral, is now called—again, unjustly—a *Lambert quadrilateral*, after Johann Heinrich Lambert (1728–1777), who used it for the same purpose. We should probably call it a *semi-Thabit quadrilateral*. Thabit’s claim is that either type of Thabit quadrilateral is in fact a rectangle. If this conclusion is granted, it follows by consideration of the diagonals of a rectangle that the sum of the acute angles in a right triangle is a right angle, and this fact makes Thabit’s proof of the parallel postulate work.

The argument of Thabit ibn-Qurra, according to Gray, is illustrated in Fig. 27.1.¹ Given three lines l , m , and n such that l is perpendicular to n at E and m intersects it at A , making an acute angle, let W be any point on m above n and draw a perpendicular WZ from W to n . If E is between A and Z , then l must intersect m by virtue of what is now called *Pasch’s theorem*, named after Moritz Pasch (1843–1930), who stated it in 1882. This theorem asserts that a line intersecting the interior of one side of a triangle must intersect at least one other side. That much of the argument would be uncontroversial. The difficult part occurs when Z is between A and E . Thabit ibn-Qurra argued as follows. By Archimedes’ principle, some

¹We are supplementing the figure and adding steps to the argument for the sake of clarity.

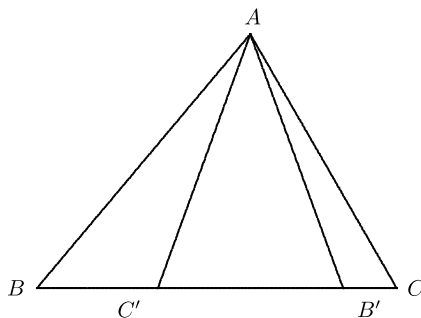


Figure 27.2. Thabit ibn-Qurra's Pythagorean theorem.

multiple of AZ , say AH , exceeds AE , so that E lies between A and H . Now by drawing a perpendicular HK to n at H , making HK equal to ZW , and joining WK , we have a Thabit quadrilateral $WZHK$, which Thabit ibn-Qurra thought he had proved to be a rectangle. Then, if X is chosen so that $AW = WX$ and a perpendicular XU is drawn to WK , the triangles AWZ and WXU will be congruent because $\angle ZWU$ is a right angle, and the sum of the acute angles of a right triangle is a right angle. Thus, because angles AWZ and WXU are both complementary to angle XWU , they are equal. It follows that $\triangle AZW \simeq \triangle WUX$ by the angle-angle-side criterion. Then WU will equal AZ . We can then start over, since WK will be less than AH by a length equal to AZ . In this way, in a finite number of steps, we will reach a point N on line m that is also on the extension of HK . Hence m contains points on both sides of l and therefore intersects l .

Gray has called Thabit ibn-Qurra's mistake "an interesting and deep one." It makes use of motion in geometry in a way that seems to be implied by Euclid's own arguments involving coinciding figures; that is, that they can be moved without changing their size or shape. Euclid makes this assumption in Proposition 4 of Book 1, where he "proves" the side-angle-side criterion for congruence by superposing one triangle on another. He does not speak explicitly of moving a triangle, but how else is one to imagine this superposition taking place?

Thabit ibn-Qurra also created the following generalization of the Pythagorean theorem. Consider a triangle ABC whose longest side is BC . Copy angle B with A as vertex and AC as one side, extending the other side to meet BC in point C' , and then copy angle C with A as vertex and BA as one side, extending the other side to meet BC in point B' , so that angle $AB'B$ and angle $AC'C$ both equal angle A . It then follows that the triangles $B'AB$ and CAC' are both similar to the original triangle ABC , and so $\overline{AB}^2 = \overline{BC} \cdot \overline{BB'}$ and $\overline{AC}^2 = \overline{BC} \cdot \overline{CC'}$, hence

$$\overline{AB}^2 + \overline{AC}^2 = \overline{BC}(\overline{BB'} + \overline{CC'}).$$

The case when angle A is acute is shown in Fig. 27.2.

27.3. AL-BIRUNI: TRIGONOMETRY

The Islamic mathematicians became familiar with both the chord tables of Ptolemy and the sine tables of Aryabhata I. They used both in their work, but it was the sine function

that they developed most fully, eventually creating all six of the ratios that we now call trigonometric functions (although these functions were lines rather than ratios to them). For the sine function, they took over the Sanskrit term *jya*, meaning *bowstring*, in its variant form *jiva*, and wrote it in the Arabic alphabet, without vowels, as *j-y-b*. This foreign word eventually became conflated with an Arabic word *jayb*, which means the pocket in a garment. According to Plofker (2009, p. 257), al-Biruni wrote that the Hindus

... call the half-Chords *juyüb* [plural of *jayb*], for the name of the Chord in the Indian [language] was *jiba*...

It was this Arabic word that eventually came to be translated into Latin as *sinus*, a word that also means a pocket or cavity.²

27.4. AL-KUHI

A mathematician who devoted himself almost entirely to geometry was Abu Sahl al-Kuhi (ca. 940–ca. 1000), the author of many works, of which some 30 survive today. Berggren (1989), who has edited these manuscripts, notes that 14 of them deal with problems inspired by the reading of Euclid, Archimedes, and Apollonius, while 11 others are devoted to problems involving the compass, spherical trigonometry, and the theory of the astrolabe. Berggren presents as an example of al-Kuhi's work the angle trisection shown in Fig. 27.3. In that figure the angle φ to be trisected is ABG , with the base BG horizontal. The idea of the trisection is to extend side AB any convenient distance to D . At the midpoint of BD , draw a pair of mutually perpendicular lines, one of which makes an angle with the horizontal equal to $\varphi/2$. Next, draw the rectangular hyperbola through B having those lines as asymptotes. Then BE is drawn equal to BD . That is, a circle through D with center at B is drawn, and its intersection with the hyperbola is labeled E . Finally, EZ is drawn parallel to BG . It then follows that $\varphi = \angle AZE = \angle ZBE + \angle ZEB = 3\theta$, as required. (The difficult part of this proof lies in showing that $\angle ZEB = \angle BDE$, as marked in Fig. 27.3.)

27.5. AL-HAYTHAM AND IBN-SAHL

Abu Ali ibn al-Haytham, known in the West as Alhazen, was the author of more than 90 books, 55 of which survive.³ His mathematical prowess is shown by his ambitious attempt to reconstruct the lost Book 8 of Apollonius' *Conics*. His most famous book is his *Treatise on Optics* (*Kitab al-Manazir*) in seven volumes. The fifth volume contains the problem known as Alhazen's problem: *Given the location of a surface, an object, and an observer, find the point on the surface at which a light ray from the object will be reflected to the observer*. Rashed (1990) points out that burning-mirror problems of this sort had been

²Contrast this Latin term with the names for the other two trigonometric functions, tangent (touching), and secant (cutting), both of which have obvious geometric meanings.

³Rashed (1989) suggested that these works and the biographical information about al-Haytham may actually refer to two different people. The opposite view was maintained by Sabra (1998).

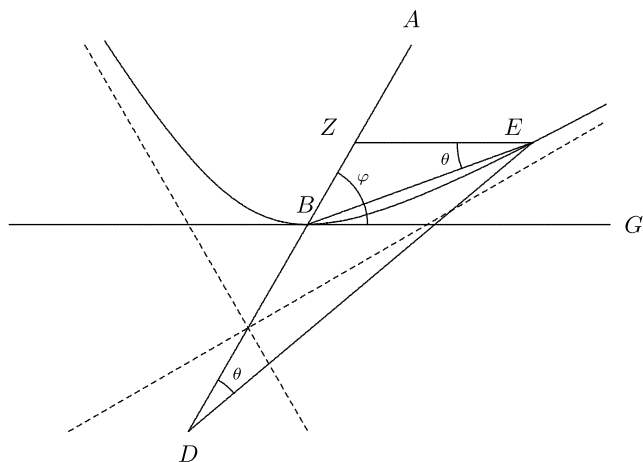


Figure 27.3. Al-Kuhi's angle trisection.

studied extensively by Muslim scholars, especially by Abu Saad ibn Sahl (940–1000) some decades before al-Haytham.

Rashed (1990, p. 478) discovered a manuscript in Teheran written by ibn Sahl containing the law of refraction known in Europe as *Snell's law*, after Willebrod Snell (1591–1626) or *Descartes' law*.⁴ The law of refraction as given by Ptolemy in the form of a table of values of the angle of refraction and the angle of incidence implied that the angle of refraction was a quadratic function of the angle of incidence. The actual relation is that the ratio of the sines of the two angles is a constant for refraction at the interface between two different media. What ibn Sahl and ibn al-Haytham knew was that the ratio of the two sines at a point where two media meet was the same whatever the angle of incidence happened to be. The seventeenth-century rediscoverers deduced theoretically that this ratio is the ratio of the speeds with which light propagates in the two media. Fermat, as we shall see in Chapter 34, showed that, given this connection, the actual time of travel from a point in one medium to a point in another is minimized.

Al-Haytham also attempted to prove the parallel postulate. According to Gray (1989, p. 45), the argument given by al-Haytham in his *Commentary on the Premises to Euclid's Book The Elements*, and later in his *Book on the Resolution of Doubts*, was based on the idea of translating a line perpendicular to a given line in such a way that it always remains perpendicular. The idea is that the endpoint of the line must trace a straight line parallel to the directing line. The idea of the proof is shown in Fig. 27.4. Al-Haytham constructs a Thabit quadrilateral $CDAE$ and imagines the side CD moving toward the opposite side EA , with D remaining on the base line, and the side remaining perpendicular at each instant of time. Obviously then the point C will remain equidistant from the base DA at all times. Al-Haytham was sure that C would move along the line CE , and could never, for example, be at a point H above that line. Unfortunately, that seemingly obvious and very intuitive conviction is precisely the point at issue in the parallel postulate.

⁴According to Guizal and Dudley, this law was stated by Thomas Harriot (1560–1621) in 1602.

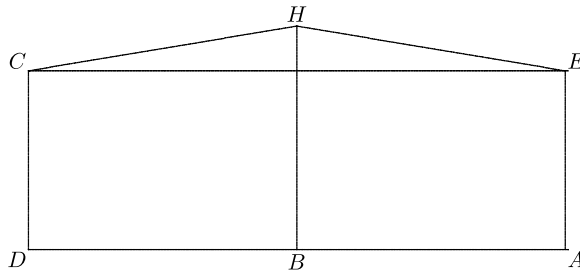


Figure 27.4. Ibn al-Haytham's attempted proof of the parallel postulate.

27.6. OMAR KHAYYAM

In his paper *Discussion of difficulties in Euclid* (Amir-Moez, 1959), Omar Khayyam raised a number of questions about al-Haytham's argument. He asked how a line could move while remaining perpendicular to a given line and, more generally, how geometry and motion could be connected. Even admitting that Euclid allowed a line to be generated by a moving point and a surface by a moving line, he pointed out that al-Haytham was requiring something more in demanding that one line remain perpendicular to another at each instant during its motion.⁵

Having refuted al-Haytham's proof, Omar Khayyam himself attempted a proof (Amir-Moez, 1959) based on a proposition that he claimed Aristotle had proved: *If two lines converge, they will (eventually) intersect*. This claim raises an interesting question, since, as we have seen, Aristotle did not accept the arguments given by scholars in Plato's Academy to prove that parallel lines exist. Given his disbelief in a completed infinity, he probably would have liked an argument proving that converging lines must intersect. Although none of the writings now attributed to Aristotle contain such an argument, Gray (1989, p. 47) suggests that Omar Khayyam may have had access to Aristotelian treatises that no longer exist. Omar Khayyam concluded on the basis of Aristotle's authority that two lines that converge on one side of a transversal must diverge on the other side. With that, having proved correctly that the perpendicular bisector of the base of a Thabit quadrilateral is also the perpendicular bisector of the summit, Omar Khayyam concluded that the base and summit could not diverge on either side, and hence must be equidistant. Like Thabit ibn-Qurra's proof, his proof depended on building one Thabit quadrilateral on top of another by doubling the common bisector of the base and summit, then crossing its endpoint with a perpendicular which (he said) would intersect the extensions of the lateral sides. Unfortunately, if that procedure is repeated often enough in hyperbolic geometry, those intersections will not occur.

All of these mathematicians were well versed in the Euclidean tradition of geometry. In the preface to his book on algebra, Omar Khayyam says that no one should attempt to read it who has not already read Euclid's *Elements* and *Data* and the first two books

⁵Omar Khayyam's objection is right on target from the point of view of modern physics. If the special theory of relativity is correct, no sense can be attached to the statement that two events occurring at different places are simultaneous. One observer may find them so, while another does not agree. The same objection applies to Thabit ibn-Qurra's argument, which assumes a rigid body. In special relativity, rigid bodies do not exist. What al-Haytham did was to ignore all points from the moving solid except those lying along a certain line.

of Apollonius' *Conics*. His reason for requiring this background was that he intended to use conic sections to solve cubic and quartic equations geometrically. This book contains Euclidean rigor attached to algebra in a way that fits equally well into the history of both algebra and geometry. In other places, it seems clear that Omar Khayyam was posing geometric problems for the sake of getting interesting equations to solve, as, for example, in the problem mentioned in the previous chapter of finding the point on a circle such that the perpendicular from the point to a radius has the same ratio to the radius that the two segments into which it divides the radius have to each other.

As his work on the parallel postulate shows, Omar Khayyam was very interested in logical niceties. In the preface to his *Algebra* and elsewhere [for example, Amir-Moez (1963, p. 328)] he shows his adherence to Euclidean standards, denying the reality of a fourth dimension:

If the algebraist were to use the square of the square in measuring areas, his result would be figurative [theoretical] and not real, because it is impossible to consider the square of the square as a magnitude of a measurable nature. . . This is even more true in the case of higher powers. [Kasir, 1931, p. 48]

27.7. NASIR AL-DIN AL-TUSI

The thirteenth century was disruptive to the Islamic world. This was the time of the Mongol expansion, which brought the conquest of China in the early part of the century, then the conquest of Kievan Rus in 1243, and, finally, the sack of Baghdad in 1258. Despite the turbulent times, the astronomer–mathematician Nasir al-Din al-Tusi (1201–1274) managed to produce some very good mathematics. Al-Tusi was treated with respect by the Mongol conqueror of Baghdad, who even built for him an astronomical observatory, at which he made years of accurate observations and improved the models in Ptolemy's *Almagest*. Al-Tusi continued the Muslim work on the problem of the parallel postulate. According to Gray (1989, pp. 50–51), al-Tusi's proof followed the route of proving that the summit angles of a Thabit quadrilateral are right angles. He showed by arguments that Euclid would have accepted that they cannot be obtuse angles, since, if they were, the summit would diverge from the base as a point moves from either summit vertex toward the other. Similarly, he claimed, they could not be acute, since in that case the summit would converge toward the base as a point moves from either summit vertex toward the other. Having thus argued that a Thabit quadrilateral must be a rectangle, he could give a proof similar to that of Thabit ibn-Qurra to establish the parallel postulate.

In a treatise on quadrilaterals written in 1260, al-Tusi also reworked the trigonometry inherited from the Greeks and Hindus and developed by his predecessors in the Muslim world, including all six triangle ratios that we know today as the trigonometric functions. In particular, he gave the law of sines for spherical triangles, which states that the sines of great-circle arcs forming a spherical triangle are proportional to the sines of their opposite angles. According to Hairetdinova (1986), trigonometry had been developing in the Muslim world for some centuries before this time, and in fact the mathematician Abu Abdullah al-Jayyani (989–1079), who lived in the Caliphate of Cordoba, wrote *The Book on Unknown Arcs of a Sphere*, a treatise on plane and spherical trigonometry. Significantly, he treated ratios of lines as numbers, in accordance with the evolution of thought on this subject in the Muslim world. Like other Muslim mathematicians, though, he does not use

negative numbers. As Hairetdinova mentions, there is evidence of Muslim influence in the first trigonometry treatise written by Europeans, the book *De triangulis omnimodis* by Regiomontanus, (Johann Müller, 1436–1476) whose exposition of plane trigonometry closely follows that of al-Jayyani.

Among these and many other discoveries, al-Tusi discovered the interesting theorem that if a circle rolls without slipping inside a circle twice as large, each point on the smaller circle moves back and forth along a diameter of the larger circle. This fact is easy to prove and an interesting exercise in geometry. It has obvious applications in geometric astronomy, and was rediscovered three centuries later by Nicolaus Copernicus (1473–1543) and used in Book 3, Chapter 4 of his *De revolutionibus*.

PROBLEMS AND QUESTIONS

Mathematical Problems

- 27.1.** Explain how Thabit ibn-Qurra's generalization of the Pythagorean theorem reduces to that theorem when angle A is a right angle. What does the figure look like if angle A is obtuse? Is there an analogous theorem if BC is not the longest side of the triangle?
- 27.2.** Al-Haytham's attempted proof of the parallel postulate is fallacious because in non-Euclidean geometry two straight lines cannot be equidistant at all points. Thus in a non-Euclidean space the two rails of a railroad cannot both be straight lines. Assuming Newton's laws of motion (an object that does not move in a straight line must be subject to some force), show that in a non-Euclidean universe one of the wheels in a pair of opposite wheels on a train must be subject to some unbalanced force at all times. [Note: The spherical earth that we live on happens to be non-Euclidean. Therefore a pair of opposite wheels on a train cannot both be moving in a great circle on the earth's surface at any time.]
- 27.3.** A two-part question: (1) Prove that in both hyperbolic and Euclidean geometry, if a line passes through the midpoint of side AB of triangle ABC and is perpendicular to the perpendicular bisector of the side BC , then it also passes through the midpoint of AC . [Hint: This is easier than it looks: Consider the line that *does* pass through both midpoints, and show that it is perpendicular to the perpendicular bisector of BC ; then argue that there is only one line passing through the midpoint of AB that is perpendicular to the perpendicular bisector of BC .] (2) Use the previous result to prove, independently of the parallel postulate, that the line joining the midpoints of the lateral sides of a Thabit (Saccheri) quadrilateral bisects both diagonals. (In Euclidean geometry, where a Thabit quadrilateral is a rectangle, the diagonals bisect each other; this is not the case in non-Euclidean geometries.)

Historical Questions

- 27.4.** What efforts were made by the Islamic geometers to clarify the theory of parallel lines and the parallel postulate of Euclid?
- 27.5.** What generalization of the Pythagorean theorem is due to Thabit ibn-Qurra?
- 27.6.** What advances in trigonometry are due to Nasir al-Din al-Tusi?

Questions for Reflection

- 27.7.** Why was speculation on the theory of parallel lines confined to the Hellenistic and Islamic geometers? Why was this problem never addressed by the Indian, Chinese, or Japanese mathematicians?
- 27.8.** Why was al-Haytham's attempt to prove the parallel postulate fallacious?
- 27.9.** What applications can you find for Nasir al-Tusi's theorem about a circle rolling without slipping inside a circle whose radius equals the diameter of the inner circle? (Imagine the circles roughened so has to have gear teeth of the same size that mesh. What use could you make of such a linkage?)

EUROPEAN MATHEMATICS, 500–1900

The background to modern mathematics lies in the Medieval period in Europe, when scholars assimilated the knowledge of the Islamic world and recovered some of the Greek works. By the fourteenth century, European mathematicians were beginning to contribute new ideas of fundamental importance, such as the representation of variable quantities on a coordinate system. In the next seven chapters, we shall trace this complicated development through the Medieval and Renaissance periods, ending around the year 1900. By that time, ideas that had been used individually for centuries had been combined in new ways to produce the calculus, which was then applied to study an immense variety of physical phenomena. Our treatment of the eighteenth and nineteenth centuries is skewed toward the calculus and its outgrowths. Other topics developed during this period, such as probability and non-Euclidean geometry, will be discussed in Part VII, which consists of surveys of some areas of mathematics in the modern era.

Contents of Part VI

This part of our history will bring the story of mathematics up just past its greatest watershed: the seventeenth-century development of calculus and its extensive use in applications during the eighteenth. It consists of the following seven chapters.

1. Chapter 28 (Medieval and Modern Europe, 500–1900) situates the mathematics developed during this period in the context of European history in general, giving some details of what was preserved from the Roman Empire, what was acquired from the Islamic world, and what the Europeans made of this heritage.
2. Chapter 29 (European Mathematics, 1200–1500) discusses European mathematical innovations during the later Medieval period.
3. Chapter 30 (Sixteenth-Century Algebra) focuses on the solution of cubic and quartic equations in Italy, the consolidation of those advances through improved notation, and the development of logarithms.
4. Chapter 31 (Renaissance Art and Geometry) takes up the topic of projective geometry in relation to the work of artists of the time.

5. Chapter 32 (The Calculus Before Newton and Leibniz) traces the development of algebra and the incorporation of infinitesimal methods into it during the early seventeenth century, a process that revealed the essential core of calculus in an unsystematic manner.
6. Chapter 33 (Newton and Leibniz) discusses the brilliant synthesis of algebra and infinitesimal methods in the work of Newton and Leibniz and their disciples.
7. Chapter 34 (Consolidation of the Calculus) is devoted to the new areas of mathematics generated by the calculus, such as differential equations and calculus of variations, along with the philosophical and foundational issues raised by admitting infinitesimal methods into mathematics.

Medieval and Early Modern Europe

Greek mathematics held on longer in the Byzantine Empire than in Western Europe. Although Theon of Alexandria had found it necessary to water down the more difficult parts of Greek geometry for the sake of his weak students, the degeneration in Latin works was even greater. The decline of cities in the West as the authority of the Roman Emperor failed was accompanied by a decline in scholarship. Only in the monasteries was learning preserved. As a result, documents from this period tend to be biased toward issues that concern the clergy.

28.1. FROM THE FALL OF ROME TO THE YEAR 1200

During the first five centuries after the fall of Rome in 476, a great deal of scholarly work was lost. While a new, and in many ways admirable, medieval civilization was being built up, only some very basic mathematics was being preserved in Western Europe. However, within the Carolingian Empire, the foundation for more advanced activity was being laid in the cathedral and monastery schools, so that when the knowledge achieved in the Islamic world was translated into Latin, scholars were prepared to appreciate and extend it. We shall mention only a handful of the scholars from this time.

28.1.1. Boethius and the Quadrivium

The philosopher Boethius (480–524) wrote Latin translations of many classical Greek works of mathematics and philosophy. His works on mathematics were translations based on Nicomachus and Euclid. Boethius' translation of Euclid has been lost. However, it is believed to be the basis of many other medieval manuscripts, some of which use his name. These are referred to as “Boethius” or pseudo-Boethius. The works of “Boethius” fit into the classical quadrivium of arithmetic, geometry, music, and astronomy. This quadrivium (fourfold path) was neatly subdivided into the categories of number (discrete quantity), magnitude (continuous quantity), statics, and kinematics. Thus number at rest is arithmetic, number in motion is music, magnitude at rest is geometry, magnitude in motion is astronomy.

Politically and militarily, the fifth century was full of disasters in Italy, and some of the best minds of the time turned from public affairs to theological questions. For many of



The quadrivium. From left to right: Music holding an instrument, arithmetic doing a finger computation, geometry studying a set of diagrams, astrology holding a set of charts. Copyright © Foto Marburg/Art Resource.

these thinkers, mathematics came to be valued especially because it could inspire religious feelings. The pseudo-Boethius gives a good example of this point of view. He writes¹:

The utility of geometry is threefold: for work, for health, and for the soul. For work, as in the case of a mechanic or architect; for health, as in the case of the physician; for the soul, as in the case of the philosopher. If we pursue this art with a calm mind and diligence, it is clear in advance that it will illuminate our senses with great clarity and, more than that, will show what it means to subordinate the heavens to the soul, to make accessible all the supernal mechanism that cannot be investigated by reason in any other way and through the sublimity of the mind beholding it, also to integrate and recognize the Creator of the world, who veiled so many deep secrets.

28.1.2. Arithmetic and Geometry

Besides the geometry just mentioned, Boethius also discussed the numerical part of the quadrivium, including a topic that is not in the older Greek works: the abacus. It was a ruled board, not the device we now call an abacus. The Latin word originally denoted the square

¹This quotation can be read online at <http://pld.chadwyck.com>. This passage is from Vol. 63. It can be reached by searching under “geometria” as title.

stone at the top of a pillar. This computational aspect of arithmetic is not so well represented in the Greek texts. In terms of its number-theoretic content, however, Boethius' treatise is far less sophisticated than the elaborate logical system found in Books VII–IX of Euclid's *Elements*.

28.1.3. Music and Astronomy

The other two sections of Boethius' work on the quadrivium are also derivative and based on Greek sources. His astronomy omits all the harder parts of Ptolemy's treatise. In addition, he wrote an influential book with the title *De institutione musica* that is of interest in the history of mathematics, since it adopts the traditional Platonic (Pythagorean) point of view that music is a subdivision of arithmetic. Boethius divides the subject of music into three areas: *Musica Mundana*, which encompasses the "music of the spheres," that is, the regular mathematical relations observed in the stars and reflected in the sounds of nature; *Musica Humana*, which reflects the orderliness of the human body and soul; and *Musica Instrumentalis*, the music produced by physical instruments, which exemplify the principles of order that the Pythagoreans allegedly ascribed to musical instruments, particularly in the simple mathematical relations between pitch and length of a string.

For over a millennium, such ideas had a firm grasp on writers such as Dante and scientists such as the seventeenth-century mathematician and astronomer Johannes Kepler. Indeed, *De institutione musica* was used as a textbook at Oxford until the eighteenth century, and Kepler actually *wrote* the music of the spheres as he conceived it.

28.1.4. The Carolingian Empire

From the sixth to the ninth centuries a considerable amount of classical learning was preserved in the monasteries in Ireland, which had been spared some of the tumult that accompanied the decline of Roman power in the rest of Europe. From this source came a few scholars to the court of Charlemagne to teach Greek and the quadrivium during the early ninth century. Charlemagne's attempt to promote the liberal arts, however, encountered great obstacles, as his empire was divided among his three sons after his death. In addition, the ninth and tenth centuries saw the last waves of invaders from the north—the Vikings, who disrupted commerce and civilization both on the continent and in Britain and Ireland until they themselves became Christians and adopted a settled way of life. Nevertheless, Charlemagne's directive to create cathedral and monastery schools had a permanent effect, leading eventually the synthesis of observation and logic known as modern science.

28.1.5. Gerbert

In the chaos that accompanied the breakup of the Carolingian Empire and the Viking invasions, the main source of stability was the Church. A career in public life for one not of noble birth was usually an ecclesiastical career, and church officials had to play both pastoral and diplomatic roles. That some of them also found time for scholarly activity is evidence of remarkable talent.

Such a talent was Gerbert of Aurillac. He was born to lower-class but free parents in south-central France some time in the 940s. He benefited from Charlemagne's decree

that monasteries and cathedrals must have schools and was educated in Latin grammar at the monastery of St. Gerald in Aurillac. Throughout a vigorous career in the Church that led to his coronation as Pope Sylvester II² in the year 999, he worked for a revival of learning, both literary and scientific. His work as secretary to the Archbishop of Reims was reported by a monk of that city named Richer, who described an abacus constructed to Gerbert's specifications. It was said to have been divided into 27 parts, and Gerbert astounded audiences with his skill in multiplying and dividing large numbers on this device (Lattin, 1961, p. 46).

While revising the curriculum in arithmetic, Gerbert wrote a tract on the use of the abacus in which the Hindu–Arabic numerals were used. This innovation required reintroduction several times, but received a strong impetus two centuries later from the *Liber abaci* of Leonardo of Pisa.

In some early letters written addressed to the monk Constantine of Fleury just before he became Abbot of Bobbio, Gerbert discusses some passages in Boethius' *Arithmetic*; and in the last letter written before he became pope, he writes to Adalbold of Liège about an inconsistency in Boethius' work (Lattin, 1961). He discusses an equilateral triangle of side 30 and height 26 (since $26 \approx 15\sqrt{3}$), whose area is therefore 390. He says that if the triangle is measured by the arithmetical rule given by Boethius—that is, in terms of its side only—the rule is “one side is multiplied by the other and the number of one side is added to this multiplication, and from this sum one-half is taken.” In our terms this would give area $s(s + 1)/2$ to an equilateral triangle of side s . We recognize here the formula for a triangular number. Thus, guided by arithmetical considerations and triangular numbers, one would expect that this formula should give the correct area. However, in the case being considered, the rule leads to an area of 465, which is too large by 20%. Gerbert correctly deduces that Boethius' rule actually gives the area of a cross section of a stack of rectangles containing the triangle in question and that the excess results from the pieces of the rectangles sticking outside the triangle. He includes a figure to explain this point to Adalbold.

We can see from this discussion by one of the leading scholars of Europe regarding the extent to which scientific and mathematical knowledge had sunk to an elementary level a thousand years ago. From these humble beginnings, European knowledge of science underwent an amazing growth over the next few centuries.

Gerbert also wrote a treatise on geometry based on Boethius. His reasons for studying geometry were similar to those given by Boethius³:

Indeed the utility of this discipline to all lovers of wisdom is the greatest possible. For it leads to vigorous exercises of the soul, and the most subtle demands on the intuition, and to many certain inquiries by true reasoning, in which wonderful and unexpected and joyful things are revealed to many along with the wonderful vigor of nature, and to contemplating, admiring, and praising the power and ineffable wisdom of the Creator who apportioned all things according to number and measure and weight; it is replete with subtle speculations.

²He was not a successful clergyman or pope. He got involved in the politics of his day, offended the Emperor, and was suspended from his duties as Archbishop of Reims by Pope Gregory V in 998. He was installed as pope by the 18-year-old Emperor Otto III, but after only three years both he and Otto were driven from Rome by a rebellion. Otto died trying to reclaim Rome, and Sylvester II died shortly afterward.

³This quotation can be read online at <http://pld.chadwyck.com>. This passage is from Vol. 139. It can be reached by searching under “geometria” as title.

This view of geometry was to be echoed four centuries later in the last Canto of Dante's *Divine Comedy*, which makes use of geometric analogs to describe the poet's vision of heaven:

Like the geometer who applies all his powers
To measure the circle, but does not find
By thinking the principle he needs,

Such was I, in this new vista.
I wished to see how the image came together
With the circle and how it could be divined there.

But my own wings could not have made the flight
Had not my mind been struck
By a flash in which his will came to me.

In this lofty vision I could do nothing.
But now turning my desire and will,
Like a wheel that is uniformly moved,

Was the love that moves the sun and the other stars.

28.1.6. Early Medieval Geometry

A picture of the level of geometric knowledge in the eleventh and twelfth centuries, before there was any major influx of translations of Arabic and Greek treatises, can be gained from an early twelfth-century treatise called *Practica geometriae* (*The Practice of Geometry*, Homann, 1991), attributed to Master Hugh of the Abbey of St. Victor in Paris.

The content of the *Practica geometriae* is aimed at the needs of surveying and astronomy and resembles the treatise of Gerbert in its content. This geometry, although elementary, is by no means unsophisticated. It discusses similar triangles and spherical triangles, using three mutually perpendicular great circles to determine positions on the sphere. After a discussion of the virtues and uses of the astrolabe, the author takes up the subjects of "altimetry" (surveying) and "cosmimetry" (astronomical measurements).

The discussion of "altimetry" is a straightforward application of similar triangles to measure inaccessible distances. The section on "cosmimetry" is of interest for two reasons. First, it gives a glimpse of what was remembered of ancient work in this area; and second, it shows what techniques were used for astronomical measurements in the twelfth century. The author begins by giving the history of measurements of the diameter of the earth, saying that the earth seems large to us, due to our confinement to its surface, even though "Compared to the incomprehensible immensity of the celestial sphere with everything in its ambit, earth, one must admit, seems but an indivisible point."

These views had been expressed by Ptolemy as justification for idealizing the earth as a point in his astronomy, and, of course, they are completely in accord with modern knowledge of the size of the cosmos. The author then goes on to discuss in detail the history of measurements of the circumference of the earth. He tells the famous story of Eratosthenes' measurement of a degree of latitude. (See Section 1 of Chapter 17.)

The author of the *Practica geometriae* continues by calculating the height of the sun by use of similar triangles. To do this, one must know the distance from the point of measurement to the point where the sun is directly overhead and then measure the length of the noontime shadow cast by a pole of known height. The author says that the Egyptians

should be given credit as the first to compute solar altitude this way and that they were successful because their country was flat and close to the sun! The figure cited for the diameter of the sun's orbit (this is geocentric astronomy) is $9,720,181 + \frac{1}{2} + \frac{7}{22}$ miles. Using the value $\pi = \frac{22}{7}$, the author computes the length of the sun's orbit as $30,549,142 \frac{5}{6} + \frac{1}{42}$ miles. (This number is less than 6% of the true value.)

28.1.7. The Translators

During the twelfth and thirteenth centuries, European scholars sought out and translated works from Arabic and ancient Greek into Latin. We can list just a few of the translators and their works here. Our debt to these people is enormous, as they greatly increased the breadth and depth of knowledge of natural science and mathematics in Europe.

1. *Adelard of Bath* (ca. 1080–1160). Born in Bath, England, Adelard (or Athelhard) studied at Tours, France in one of the cathedral schools established by Charlemagne, as Gerbert had done. He traveled widely throughout the Mediterranean region. Some time in the second decade of the twelfth century, he translated Euclid's *Elements* into Latin from an Arabic manuscript. This translation became the basis for all Latin translations of this work for the next few centuries. He also translated al-Khwarizmi's astronomical tables, the Arabic original of which no longer exists.
2. *Plato of Tivoli* (early twelfth century). Little is known of the life of Plato Tiburtinus (Plato of Tivoli). He is best known for translating al-Battani's *Kitab al-Zij* (*Book of Astronomy*) into Latin as *De motu stellarum*.
3. *Robert of Chester* (twelfth century). Robert of Chester was an Englishman who went to Segovia, Spain. He translated al-Khwarizmi's *Algebra* around 1145.
4. *Gherard of Cremona* (1114–1187). Born in Cremona, Italy, Gherard traveled to Spain with the intention of studying the works of Ptolemy. He made translations of some eighty works from Arabic into Latin, including an edition of the *Elements* edited by Thabit ibn-Qurra, al-Khwarizmi's *Algebra*, and of course the *Almagest*.

Various authors ascribe to each of these last three translators the responsibility for translating the Arabic word *jayb*, which had evolved from the Sanskrit *jiva* (bowstring), into Latin as *sinus*, thereby establishing a usage that has persisted for 900 years all over Europe. Most of these statements are vague as to precisely where the term occurs. According to Holt, Lambton, and Lewis (1970, p. 754), the word occurs first in the twelfth-century translation of al-Battani's *Kitab al-Zij*, and therefore must have been introduced by Plato of Tivoli.

28.2. THE HIGH MIDDLE AGES

As the western part of the world of Islam was growing politically and militarily weaker because of invasion and conquest, Europe was entering on a period of increasing power and vigor. One expression of that new vigor, the stream of European mathematical creativity that began as a small rivulet 1000 years ago, has been steadily increasing until now; it is an enormous river and shows no sign of subsiding. By the middle of the twelfth century, European civilization had absorbed much of the learning of the Islamic world and was ready to embark on its own explorations. This was the zenith of papal power in Europe, exemplified by the ascendancy of the popes Gregory VII (1073–1085) and Innocent III

(1198–1216) over the emperors and kings of the time. The Emperor Frederick I, known as Frederick Barbarossa because of his red beard, ruled the empire from 1152 to 1190 and tried to maintain the principle that his power was not dependent on the Pope, but was ultimately unsuccessful. His grandson Frederick II (1194–1250) was a cultured man who encouraged the arts and sciences. To his court in Sicily⁴ he invited distinguished scholars of many different religions, and he corresponded with many others. He himself wrote a treatise on the principles of falconry. He was in conflict with the Pope for much of his life and even tried to establish a new religion, based on the premise that “no man should believe aught but what may be proved by the power and reason of nature,” as the papal document excommunicating him stated.

Our list of memorable European mathematicians from the late Medieval period begins in the empire of Frederick II.

28.2.1. Leonardo of Pisa

Leonardo (1170–1250) says in the introduction to his major book, the *Liber abaci*, that he accompanied his father on an extended commercial mission in Algeria with a group of Pisan merchants. There, he says, his father had him instructed in the Hindu–Arabic numerals and computation, which he enjoyed so much that he continued his studies while on business trips to Egypt, Syria, Greece, Sicily, and Provence. Upon his return to Pisa he wrote a treatise to introduce this new learning to Italy. The treatise, whose author is given as “Leonardus filius Bonaccij Pisani,” that is, “Leonardo, son of Bonaccio of Pisa,” bears the date 1202. In the nineteenth century Leonardo’s works were edited by the Italian nobleman Baldassare Boncompagni (1821–1894), who also compiled a catalog of locations of the manuscripts (Boncompagni, 1854). The name Fibonacci by which the author is now known seems to have become generally used only in the nineteenth century. A history of what is known of Leonardo’s life and an exposition of his mathematical works has recently appeared (Devlin, 2011).

28.2.2. Jordanus Nemorarius

The works of Archimedes were translated into Latin in the thirteenth century, and his work on the principles of mechanics was extended. One of the authors involved in this work was Jordanus Nemorarius (1225–1260). Little is known about this author except certain books that he wrote on mathematics and statics for which manuscripts still exist dating to the actual time of composition. One of his works, *Liber Jordani de Nemore de ratione ponderis* [*The book of Jordanus Nemorarius on the ratio of weight* (Claggett, 1960, pp. 167–229)] contains the first correct statement of the mechanics of an inclined plane. We shall confine our discussion, however, to his algebraic work, in which he discussed various conditions from which the explicit value of a number can be deduced.

28.2.3. Nicole d’Oresme

One of the most distinguished of the medieval philosophers was Nicole d’Oresme (1323–1382), whose clerical career brought him to the office of Bishop of Lisieux in 1377.

⁴Sicily was reconquered from the Muslims in the eleventh century by the Normans. Being in contact with all three of the great Mediterranean civilizations of the time, it was the most cosmopolitan center of culture in the world for the next two centuries.

D'Oresme had a wide-ranging intellect and studied economics, physics, and mathematics as well as theology and philosophy. He considered the motion of physical bodies from various points of view, formulated the Merton rule of uniformly accelerated motion (named for Merton College, Oxford), and for the first time in history explicitly used one line to represent time, a line perpendicular to it to represent velocity, and the area under the graph (as we would call it) to represent distance.

28.2.4. Regiomontanus

The work of translating the Greek and Arabic mathematical works went on for several centuries. One of the last to work on this project was Johann Müller (1436–1476) of Königsberg, better known by his Latin name of Regiomontanus, a translation of Königsberg (King's Mountain). Although he died young, Regiomontanus made valuable contributions to astronomy, mathematics, and the construction of scientific measuring instruments. He studied in Leipzig while a teenager and then spent a decade in Vienna and the decade following in Italy and Hungary. The last five years of his life were spent in Nürnberg. He is said to have died of an epidemic while in Rome as a consultant to the Pope on the reform of the calendar.

Regiomontanus checked the data in copies of Ptolemy's *Almagest* and made new observations with his own instruments. He laid down a challenge to astronomy, remarking that further improvement in theoretical astronomy, especially the theory of planetary motion, would require more accurate measuring instruments. He established his own printing press in Nürnberg so that he could publish his works. These works included several treatises on pure mathematics. He established trigonometry as an independent branch of mathematics rather than a tool in astronomy. The main results we now know as plane and spherical trigonometry are in his book *De triangulis omnimodis*, although not exactly in the language we now use.

28.2.5. Nicolas Chuquet

The French Bibliothèque Nationale is in possession of the original manuscript of a mathematical treatise written at Lyons in 1484 by one Nicolas Chuquet (1445–1488). Little is known about the author, except that he describes himself as a Parisian and a man possessing the degree of Bachelor of Medicine. The treatise (see Flegg, 1988) consists of four parts: a treatise on arithmetic and algebra called *Triparty en la science des nombres*, a book of problems to illustrate and accompany the principles of the *Triparty*, a book on geometrical measurement, and a book of commercial arithmetic. The last two are applications of the principles in the first book.

28.2.6. Luca Pacioli

Written at almost the same time as Chuquet's *Triparty* was a work called the *Summa de arithmetica, geometrica, proportioni et proportionalita* by Luca Pacioli (or Paciuolo, 1445–1517). Since Chuquet's work was not printed until the nineteenth century, Pacioli's work is believed to be the first Western printed work on algebra. In comparison with the *Triparty*, however, the *Summa* seems less original. Pacioli has only a few abbreviations, such as *co* for *cosa*, meaning *thing* (the unknown), *ce* for *censo* (the square of the unknown), and *æ* for *æquitur* (equals). Despite its inferiority to the *Triparty* where symbolism is concerned,

the *Summa* was much the more influential of the two books, because it was published. It is referred to by the Italian algebraists of the early sixteenth century as a basic source.

28.2.7. Leon Battista Alberti

In art, the fifteenth century was a period of innovation that marked the beginning of the period we call the Renaissance. In an effort to give the illusion of depth in two-dimensional representations, some artists looked at geometry from a new point of view, studying the projection of two- and three-dimensional shapes in two dimensions to see what properties were preserved and how others were changed. A description of such a procedure, based partly on the work of his predecessors, was given by Leon Battista Alberti (1404–1472) in a 1435 Latin treatise entitled *De pictura*, published posthumously in Italian as *Della pittura* in 1511.

28.3. THE EARLY MODERN PERIOD

Sixteenth-century Italy produced a group of sometimes quarrelsome but always brilliant algebraists, who worked to advance mathematics in order to achieve academic success and for the pleasure of discovery. As happened in Japan a century later, each new advance brought a challenge for further progress.

28.3.1. Scipione del Ferro

A method of solving a particular cubic equation was discovered by a lector (reader, that is, a tutor) at the University of Bologna, Scipione del Ferro (1465–1525), around the year 1500.⁵ He communicated this discovery to another mathematician, Antonio Maria Fior (dates unknown), who then used the knowledge to win mathematical contests.

28.3.2. Niccolò Tartaglia

Fior met his match in 1535, when he challenged Niccolò Fontana of Brescia, (1500–1557) known as Tartaglia (the Stammerer) because a wound he received as a child when the French overran Brescia in 1512 left him with a speech impediment. Tartaglia had also discovered how to solve certain cubic equations and thus won the contest.

28.3.3. Girolamo Cardano

A brilliant mathematician and gambler, who became rector of the University of Padua at the age of 25, Girolamo Cardano (1501–1576) was writing a book on mathematics in 1535 when he heard of Tartaglia's victory over Fior. He wrote to Tartaglia asking permission to include this technique in his work. Tartaglia at first refused, hoping to work out all the

⁵Before modern notation was introduced, there was no uniform way of writing a general cubic equation. Since negative numbers were not understood, equations had to be classified according to the terms on each side of the equality. As we saw in the case of Omar Khayyam, this complication results in many different types of cubics, each requiring a special algorithm for its solution.

details of all cases of the cubic and write a treatise himself. According to his own account, Tartaglia confided the secret of one kind of cubic to Cardano in 1539, after Cardano swore a solemn oath not to publish it without permission and gave Tartaglia a letter of introduction to the Marchese of Vigevano. Tartaglia revealed a rhyme by which he had memorized the procedure.

Tartaglia did not claim to have given Cardano any proof that his procedure works. It was left to Cardano himself to find the demonstration. Cardano kept his promise not to publish this result until 1545. However, as Tartaglia delayed his own publication, and in the meantime Cardano had discovered the solution of other cases of the cubic himself and had also heard that del Ferro had priority anyway, he published the result in his *Ars magna* (*The Great Art*), giving credit to Tartaglia. Tartaglia was furious and started a bitter controversy over Cardano's alleged breach of faith.

28.3.4. Ludovico Ferrari

Cardano's student Ludovico Ferrari (1522–1565) worked with him in the solution of the cubic, and between them they had soon found a way of solving certain quartic equations.

28.3.5. Rafael Bombelli

In addition to the mathematicians proper, we must also mention an engineer in the service of an Italian nobleman. Rafael Bombelli (1526–1572) is the author of a treatise on algebra that appeared in 1572. In the introduction to this treatise we find the first mention of Diophantus in the modern era. Bombelli said that, although all authorities are agreed that the Arabs invented algebra, he, having been shown the work of Diophantus, credits the invention to the latter. In making sense of what his predecessors did, he was one of the first to consider the square root of a negative number and to formulate rules for operating with such numbers. His work in this area will be discussed in more detail in Chapter 41.

28.4. NORTHERN EUROPEAN ADVANCES

The work being done in Italy did not escape the notice of French and British scholars of the time, and important mathematical works were soon being produced in those two countries.

28.4.1. François Viète

A lawyer named François Viète (1540–1603), who worked as tutor in a wealthy family and later became an advisor to Henri de Navarre (who became the first Bourbon king, Henri IV, in 1598), found time to study Diophantus and to introduce his own ideas into algebra. His book *Artis analyticae praxis* (*The Practice of the Analytic Art*) contained some of the notational innovations that make modern algebra much less difficult than the algebra of the sixteenth century.

28.4.2. John Napier

In the late sixteenth century the problem of simplifying laborious multiplications, divisions, root extractions, and the like, was attacked by the Scottish laird John Napier, (1550–1617)

Baron of Murchiston. His work consisted of two parts, a theoretical part, based on a continuous geometric model, and a computational part, involving a discrete (tabular) approximation of the continuous model. The computational part was published in 1614. However, Napier hesitated to publish his explanation of the theoretical foundation. Only in 1619, two years after his death, did his son publish an English translation of Napier's theoretical work under the title *Mirifici logarithmorum canonis descriptio* (*A Description of the Marvelous Law of Logarithms*). This subject, although aimed at a practical end, turned out to have enormous value in theoretical studies as well.

QUESTIONS

Historical Questions

- 28.1. What mathematics was preserved in the Western part of the Roman Empire during the period from 500 to 1000?
- 28.2. What justifications do the early Medieval writers give for the study of geometry and arithmetic?
- 28.3. What Arabic and Greek works were brought into Europe in the eleventh and twelfth centuries, and who were the translators responsible for making them available in Latin?
- 28.4. How did the term *sine* (Latin *sinus*) come to have a geometric meaning as one of the trigonometric functions?

Questions for Reflection

- 28.5. Dante's final stanza (quoted above) uses the problem of squaring the circle to express the sense of an intellect overwhelmed, which was inspired by his vision of heaven. What resolution does he find for the inability of his mind to grasp the vision rationally? Would such an attitude, if widely shared, affect mathematical and scientific activity in a society?
- 28.6. What is the significance of ruling a board into 27 columns to make an abacus, as Gerbert is said to have done? Does it indicate that there was no symbol for zero?
- 28.7. One popular belief about Christopher Columbus is that he proved to a doubting public that the earth was spherical. What grounds are there for believing that "the public" doubted this fact? Which people in the Middle Ages would have been likely to believe in a flat earth? Consider also the frequently repeated story that people used to believe the stars were near the earth. Is this view of Medieval scholarship plausible in the light of the *Practica geometriae*?
- 28.8. What role can or should or does mathematics play in representational arts such as painting and sculpture? Does the presence of mathematical elements enhance or detract from the emotional content and artistic creativity involved in these arts?

European Mathematics: 1200–1500

In the previous chapter, we mentioned the emperor Frederick II, whose court was located in Sicily. His encouragement of arts and sciences gave a voice to one of the most remarkable mathematicians of the Middle Ages, Leonardo of Pisa, with whom we begin our discussion of late Medieval mathematics.

29.1. LEONARDO OF PISA (FIBONACCI)

As soon as translations from Arabic into Latin became generally available in the twelfth and thirteenth centuries, Western Europeans began to learn about algebra. The first work translated (by Robert of Chester in 1145) was al-Khwarizmi's *Algebra*. Several talented mathematicians appeared early on who were able to make original contributions to the development of algebra. In some cases the books that they wrote were not destined to be published for many centuries, but at least one of them formed part of an Italian tradition of algebra that continued for several centuries. That tradition begins with Leonardo, who wrote several mathematical works, the best known of which is the *Liber abaci*.¹

29.1.1. The *Liber abaci*

Many of the problems in the *Liber abaci* (*Book of Computation*) reflect the routine computations that must be performed when converting currencies. These are applications of the Rule of Three that we have found in Brahmagupta and Bhaskara. Many of the other problems are purely fanciful. Leonardo's indebtedness to Arabic sources was detailed by Levey (1966), who listed 29 problems in the *Liber abaci* that are identical to problems in the *Algebra* of Abu Kamil. In particular, the problem of separating the number 10 into two parts satisfying an extra condition occurs many times. For example, one problem is to find x such that $10/x + 10/(10 - x) = 6\frac{1}{4}$.

¹Devlin (2011), who has looked at the old manuscripts of this work, says that it is properly spelled *Liber abbaci*. The spelling we are using merely preserves a long-standing traditional usage.

29.1.2. The Fibonacci Sequence

The most famous (not the most profound) of Leonardo's achievements is a problem from his *Liber abaci*, whose second edition appeared in 1202: *How many pairs of rabbits can be bred from one pair in one year, given that each pair produces a new pair each month, beginning two months after its birth?*

By enumeration of cases, the author concludes that there will be 377 pairs, and "in this way you can do it for the case of infinite numbers of months." The reasoning is simple. Each month, those pairs that were alive two months earlier produce duplicates of themselves. Hence the total number of rabbits after $n + 2$ months is the number alive after $n + 1$ months plus the number alive after n months. That is, each term in the sequence is the sum of the two preceding numbers.

Assuming the original pair was a mature pair, ready to reproduce, the sequence generated in this way—starting at the beginning of the year, when 0 months have elapsed—is (1, 2, 3, 5, 8, . . .), and its 13th term is 377. This sequence has been known as the *Fibonacci sequence* since the printing of the *Liber abaci* in the nineteenth century. The Fibonacci sequence has been an inexhaustible source of identities. Many curious representations of its terms have been obtained, and there is a mathematical journal, the *Fibonacci Quarterly*, named in its honor and devoted to its lore.

A Practical Application In 1837 and 1839 the crystallographer Auguste Bravais (1811–1863) and his brother Louis (1801–1843) published articles on the growth of plants.² In these articles they studied the spiral patterns in which new branches grow out of the limbs of certain trees and classified plants into several categories according to this pattern. For one of these categories they gave the amount of rotation around the limb between successive branches as $137^\circ 30' 28''$. Now, one could hardly measure the limb of a tree so precisely. To measure within 10° would require extraordinary precision. To refine such crude measurements by averaging to the claimed precision of $1''$, that is, $1/3600$ of a degree, would require thousands of individual measurements. In fact, the measurements were carried out in a more indirect way, by counting the total number of branches after each full turn of the spiral. Many observations convinced the brothers Bravais that normally there were three branches in a little less than two turns, five in a little more three turns, eight in a little less than five turns, and thirteen in a little more than eight turns. For that reason they took the actual amount of revolution between successive branches to be the number we call $1/\Phi = (\sqrt{5} - 1)/2 = \Phi - 1$ of a complete (360°) revolution, since

$$\frac{3}{2} < \frac{8}{5} < \Phi < \frac{13}{8} < \frac{5}{3}.$$

Observe that $360^\circ \div \Phi \approx 222.4922359^\circ \approx 222^\circ 29' 32'' = 360^\circ - (137^\circ 30' 28'')$. An illustration of this kind of growth is shown in Fig. 29.1. The picture shows three views of a branch of a flowering crab apple tree with the twigs cut off and the points from which they grew marked by pushpins. When these pins are joined by string, the string follows

²See the article by I. Adler, D. Barabe, and R. V. Jean, "A history of the study of phyllotaxis," *Annals of Botany*, **80** (1997), pp. 231–244, especially p. 234. The articles by Auguste and Louis Bravais are "Essai sur la disposition générale des feuilles curvisériées," *Annales des sciences naturelles*, **7** (1837), pp. 42–110, and "Essai sur la disposition générale des feuilles rectisériées," *Congrès scientifique de France*, **6** (1839), pp. 278–330.

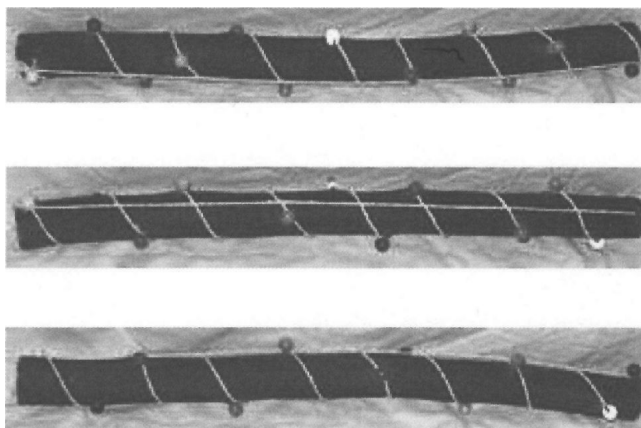


Figure 29.1. Three views of a branch of a flowering crab apple tree.

a helical path of nearly constant slope along the branch. By simply counting, one can get an idea of the *average number* of twigs per turn. For example, the fourth intersection is between pins 6 and 7, indicating that the average number of pins per turn up to that point is between $\frac{6}{4} = 1.5$ and $\frac{7}{4} = 1.75$. One can see that the pins that fall nearest to the intersection of this helical path with the meridian line marked along the length of the branch are pins numbered 3, 5, 8, and 13, which are Fibonacci numbers, and that the intersections they are near come at the end of 2, 3, 5, and 8 revolutions, respectively, also Fibonacci numbers. Thus the average number of twigs per turn is approximately $\frac{3}{2}$ or $\frac{5}{3}$ or $\frac{8}{5}$ or $\frac{13}{8}$. The brothers Bravais knew that the ratios of successive Fibonacci numbers are the terms in the continued-fraction expansion of the Golden Ratio $\Phi = (1 + \sqrt{5})/2$, and hence they chose this elegant way of formulating what they had observed. By looking at the side of the intersection where the corresponding pins are in Fig. 29.1, you can see that the first and third of these approximations are underestimates and the second and fourth are overestimates. You can also see that the approximation gets better as the number of turns increases.

This pattern is not universal among plants, although the brothers Bravais were able to find several classes of plants that exhibit a pattern of this type, with different values for the first two terms of the sequence.

29.1.3. The *Liber quadratorum*

In his *Liber quadratorum* [*Book of Squares* (Sigler, 1987)] Leonardo speculated on the difference between square and nonsquare numbers. In the prologue, addressed to the Emperor Frederick II, Leonardo says that he had been inspired to write the book because a certain John of Palermo, whom he had met at Frederick's court, had challenged him to find a square number such that if 5 is added to it or subtracted from it, the result is again a square.³ This question inspired him to reflect on the difference between square and nonsquare numbers. He then notes his pleasure on learning that Frederick had actually read one of his previous books and uses that fact as justification for writing on the challenge problem.

³Leonardo gave a general discussion of problems of this type, asking when $m^2 + kn^2$ and $m_2 + 2kn^2$ can both be squares.

The *Liber quadratorum* is written in the spirit of Diophantus and shows a keen appreciation of the conditions under which a rational number is a square. Indeed, the ninth of its 24 propositions is a problem of Diophantus: *Given a nonsquare number that is the sum of two squares, find a second pair of squares having this number as their sum.* This problem is Problem 9 of Book 2 of Diophantus, as discussed in Section 4 of Chapter 9. Leonardo's solution of this problem, like that of Diophantus, involves a great deal of arbitrariness, since the problem does not have a unique solution. The resemblance in some points is so strong that one is inclined to think that Leonardo saw a copy of Diophantus, or, more likely, an Arabic work commenting and extending the work of Diophantus. This question is discussed by the translator of the *Liber quadratorum* (Sigler, 1987, pp. xi–xii), who notes that strong resemblances have been pointed out between the *Liber quadratorum* and a book by Abu Bekr ibn Muhammad ibn al-Husayn Al-Karaji (953–1029) called the *Fakhri*,⁴ parts of which were copied from the *Arithmetica*, but that there are also parts of the *Liber quadratorum* that are original.

One advance in the *Liber quadratorum* is the use of general letters in an argument. Although in some proofs Leonardo argues much as Diophantus does, using specific numbers, he becomes more abstract in others. For example, Proposition 5 requires finding two numbers, the sum of whose squares is a square that is also the sum of the squares of two given numbers. He says to proceed as follows. Let the two given numbers be *.a.* and *.b.* and the sum of their squares *.g.* Now take any other two numbers *.de.* and *.ez.* [not proportional to the given numbers] the sum of whose squares is a square. These two numbers are arranged as the legs of a right triangle. If the square on the hypotenuse of this triangle is *.g.*, the problem is solved. If the square on the hypotenuse is larger than *.g.*, mark off the square root of *.g.* on the hypotenuse. The projections (as we would call them) of this portion of the hypotenuse on each of the legs are known, since their ratios to the square root of *.g.* are known. Moreover, that ratio is rational, since they are the same as the ratios of *.a.* and *.b.* to the hypotenuse of the original triangle. These two projections therefore provide the new pair of numbers. Being proportional to *.a.* and *.b.*, which are not proportional to the two numbers given originally, they must be different from those numbers.

This argument is more convincing, because it is more abstract, than proofs by example, but the geometric picture plays an important role in making the proof comprehensible.

29.1.4. The *Flos*

Leonardo's approach to algebra begins to look modern in other ways as well. In one of his works, called the *Flos super solutionibus quarumdam questionum ad numerum et ad geometriam vel ad utrumque pertinentum* [*The Full Development*⁵ of the Solutions of Certain Questions Pertaining to Number or Geometry or Both (Boncompagni 1854, p. 4)] he reports the challenge from John of Palermo mentioned above, which was to find a number satisfying $x^3 + 2x^2 + 10x = 20$ using the methods given by Euclid in Book 10 of the *Elements*, that is, to construct a line of this length using straightedge and compass. In working on this question, Leonardo made two important contributions to algebra, one numerical and one theoretical. The numerical contribution was to give the unique positive root in sexagesimal notation correct to six places. The theoretical contribution was to show by

⁴Apparently, this word means something like *glorious* and the full title might be translated as *The Glory of Algebra*.

⁵The word *flos* means *bloom* and can be used in the figurative sense of "the bloom of youth." That appears to be its meaning here.

using divisibility properties of numbers that there cannot be a rational solution or a solution obtained using only rational numbers and square roots of rational numbers.

29.2. HINDU–ARABIC NUMERALS

The *Liber abaci* advocated the use of the Hindu–Arabic numerals that we are familiar with. Partly because of the influence of that book, the advantages of this system came to be appreciated, and within two centuries these numerals were winning general acceptance. In 1478, an arithmetic was published in Treviso, Italy, explaining the use of Hindu–Arabic numerals and containing computations in the form shown in Fig. 26.1 of Chapter 26. In the sixteenth century, scholars such as Robert Recorde (1510–1558) in Britain and Adam Ries (1492–1559) in Germany, advocated the use of the Hindu–Arabic system and established it as a universal standard.

The system was explained by the Flemish mathematician and engineer Simon Stevin (1548–1620) in his 1585 book *De Thiende (Decimals)*. Stevin took only a few pages to explain, in essentially modern terms, how to add, subtract, multiply, and divide decimal numbers. He then showed the application of this method of computing in finding land areas and the volumes of wine vats. He wrote concisely, as he said, “because here we are writing for teachers, not students.” His notation appears slightly odd, however, since he put a circled 0 where we now have the decimal point, and thereafter he indicated the rank of each digit



From a 1535 illustration to the *Margarita philosophica (Philosophical Pearl)* published by Gregor Reisch (1467–1525) in 1503. Copyright © Foto Marburg/Art Resource.

by a similarly encircled number. For example, he would write 13.4832 as 13 ④ 4 ① 8 ② 3 ③ 2 ④ . Here is his explanation of the problem of expressing $0.07 \div 0.00004$:

When the divisor is larger [has more digits] than the dividend, we adjoin to the dividend as many zeros as desired or necessary. For example, if 7 ② is to be divided by 4 ⑤ , I place some 0s next to the 7, namely 7000. This number is then divided as above, as follows:

$$\begin{array}{r}
 \beta \ 2 \\
 7 \ 0 \ 0 \ 0 \quad (1 \ 7 \ 5 \ 0 \ ① \\
 \cancel{4} \ \cancel{4} \ \cancel{4} \ \cancel{4}
 \end{array}$$

Hence the quotient is 1750 ① (Gericke and Vogel, 1965, p. 19).

Except for the location of the digits and the cross-out marks, this notation is essentially what is now used by school children in the United States. In other countries—Russia, for example—the divisor would be written just to the right of the dividend and the quotient just below the divisor.

Stevin also knew what to do if the division does not come out even. He pointed out that when 4 ① is divided by 3 ②, the result is an infinite succession of 3s and that the exact answer will never be reached. He commented, “In such a case, one may go as far as the particular case requires and neglect the excess. It is certainly true that $13 \ ① \ 3 \ ① \ 3 \frac{1}{3} \ ②$, or $13 \ ① \ 3 \ ① \ 3 \ ② \ 3 \frac{1}{3} \ ③$, and so on, are exactly equal to the required result, but our goal is to work only with whole numbers in this decimal computation, since we have in mind what occurs in human business, where [small parts of small measures] are ignored.” Here we have a clear case in which the existence of infinite decimal expansions is admitted, without any hint of the possibility of irrational numbers. Stevin was an engineer, not a theoretical mathematician. His examples were confined to what is of practical value in business and engineering, and he made no attempt to show how to calculate with an actually infinite decimal expansion.

Stevin did, however, suggest a reform in trigonometry that was ignored until the advent of hand-held calculators, remarking that, “if we can trust our experience (with all due respect to Antiquity and thinking in terms of general usefulness), it is clear that the series of divisions by 10, not by 60, is the most efficient, at least among those that are by nature possible.” On those grounds, Stevin suggested that degrees be divided into decimal fractions rather than minutes and seconds. Modern hand-held calculators now display angles in exactly this way, despite the scornful remark of a twentieth-century mathematician that this mixture of sexagesimal and decimal notation proves that “it required four millennia to produce a system of angle measurement that is completely absurd.”

29.3. JORDANUS NEMORARIUS

The translator and editor of Jordanus’ book *De numeris datis* (*On Given Numbers*, Hughes, 1981, p. 11) says, “It is reasonable to assume...that Jordanus was influenced by al-Khwarizmi’s work.” This conclusion was reached on the basis of Jordanus’ classification of quadratic equations and his order of expounding the three types, among other resemblances between the two works.

De numeris datis is the algebraic equivalent of Euclid's *Data*. Where Euclid says that a line is given (determined) if its ratio to a given line is given, Jordanus Nemorarius says that a number is given if its ratio to a given number is given. The well-known elementary fact that two numbers can be found if their sum and difference are known is generalized to the theorem that any set of numbers can be found if the differences of the successive numbers and the sum of all the numbers is known. This book contains a large variety of data sets that determine numbers. For example, *if the sum of the squares of two numbers is known, and the square of the difference of the numbers is known, the numbers can be found*. The four books of *De numeris datis* contain about 100 such results. These results admit a purely algebraic interpretation. For example, in Book 4 Jordanus Nemorarius writes:

If a square with the addition of its root multiplied by a given number makes a given number, then the square itself will be given. [p. 100]⁶

Where earlier mathematicians would have proved this proposition with examples, Jordanus Nemorarius uses letters representing abstract numbers. The assertion is that there is only one (positive) number x such that $x^2 + \alpha x = \beta$, and that x can be found if α and β are given.

29.4. NICOLE D'ORESME

A work entitled *Tractatus de latitudinibus formarum* (*Treatise on the Latitude of Forms*) was published in Paris in 1482 and ascribed to Oresme, but probably written by one of his students. It contains descriptions of the graphical representation of “intensities.” This concept finds various expressions in physics, corresponding intuitively to the idea of density. In Oresme's language, an “intensity” is any constant of proportionality. Velocity, for example, is the “intensity” of motion.

We think of analytic geometry as the application of algebra to geometry. Its origins in Europe, however, antedate the high period of European algebra by a century or more. The first adjustment in the way mathematicians think about physical dimensions, an essential step on the way to analytic geometry, occurred in the fourteenth century. The crucial idea found in the representation of distance as the “area under the velocity curve” was that since the area of a rectangle is computed by multiplying length and width and the distance traveled at constant speed is computed by multiplying velocity and time, it follows that if one line is taken proportional to time and a line perpendicular to it is proportional to a (constant) velocity, the area of the resulting rectangle is proportional to the distance traveled.

Oresme considered three forms of qualities, which he labeled *uniform*, *uniformly difform*, and *difformly difform*. We would call these classifications constant, linear, and nonlinear. Examples are shown in Fig. 29.2, which can be found in another of Oresme's works. Oresme (or his students) realized that the “difformly difform” constituted a large class of qualities and mentioned specifically that a semicircle could be the representation of such a quality.

The advantage of representing a *distance* by an *area* rather than a line appeared in the case when the velocity changed during a motion. In the simplest nontrivial case the velocity

⁶This translation is my own and is intended to be literal; Hughes gives a smoother, more idiomatic translation on p. 168.

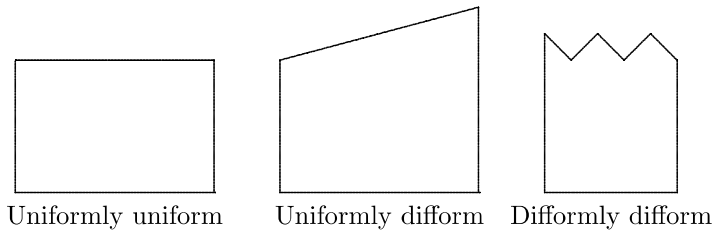


Figure 29.2. Nicole Oresme's classification of motions.

was uniformly difform. This is the case of constant acceleration. In that case, the distance traversed is what it would have been had the body moved the whole time with the velocity it had at the midpoint of the time of travel. This is the case now called *uniformly accelerated* motion. According to Clagett (1968, p. 617), this rule was first stated by William Heytesbury (ca. 1313–ca. 1372) of Merton College, Oxford around 1335 and was well known during the Middle Ages. Boyer (1949, p. 83) says that the rule was stated around this time by another fourteenth-century Oxford scholar named Richard Suiseth,⁷ known as Calculator for his book *Liber calculatorum*. Suiseth shares with Oresme the credit for having proved that the harmonic series $(1 + \frac{1}{2} + \frac{1}{3} + \dots)$ diverges.

The rule just stated is called the *Merton rule*. In his book *De configurationibus qualitatum et motuum*, Oresme applied these principles to the analysis of such motion and gave a simple geometric proof of the Merton Rule. He illustrated the three kinds of motion by drawing a figure similar to Fig. 29.2. He went on to say that if a dififormly difform quality was composed of uniform or uniformly difform parts, as in the example in Fig. 29.2, its quantity could be measured by (adding) its parts. He then pushed this principle to the limit, saying that if the quality was difform but not made up of uniformly difform parts, say being represented by a curve, then “it is necessary to have recourse to the mutual measurement of curved figures” (Clagett, 1968, p. 410). This statement must mean that the distance traveled is the “area under the velocity curve” in all three cases. Oresme unfortunately did not give any examples of the more general case, but he could hardly have done so, since the measurement of figures bounded by curves was still very primitive in his day.

29.5. TRIGONOMETRY: REGIOMONTANUS AND PITISCUS

In the late Middle Ages, the treatises translated into Latin from Arabic and Greek were made the foundation for ever more elaborate mathematical theories.

29.5.1. Regiomontanus

Analytic geometry as we know it today would be unthinkable without plane trigonometry. Latin translations of Arabic texts of trigonometry, such as the text of Nasir al-Din al-Tusi, began to circulate in Europe in the late Middle Ages. These works provided the foundation for such books as *De triangulis omnimodis* (*On General Triangles*) by Regiomontanus,

⁷Also known as Richard Swyneshed and as Swineshead with a great variety of first names. There is uncertainty whether the works ascribed to this name are all due to the same person.

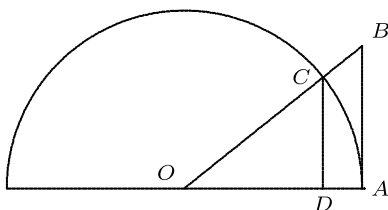


Figure 29.3. The three basic trigonometric functions: The secant OB , which cuts the circle; the tangent AB , which touches the circle; the sine CD , which is half of a chord.

published in 1533, more than half a century after the author's death. This book contained trigonometry almost in the form still taught. Book 2, for example, contains as its first theorem the law of sines for plane triangles, which asserts that the sides of triangles are proportional to the sines of the angles opposite them. The main difference between this trigonometry and ours is that a sine remains a *line* rather than a *ratio*. It is referred to an *arc* rather than to an *angle*. It was once believed that Regiomontanus discovered the law of sines for spherical triangles (Proposition 16 of Book 4) as well,⁸ but we now know that this theorem was known at least 500 years earlier to Muslim mathematicians whose work Regiomontanus must have read.

29.5.2. Pitiscus

A more advanced book on trigonometry, which reworked the reasoning of Heron on the area of a triangle given its sides, was *Trigonometriae sive de dimensione triangulorum libri quinque* (*Five Books of Trigonometry, or, On the Size of Triangles*), published in 1595 and written by the Calvinist theologian Bartholomeus Pitiscus (1561–1613). This was the book that established the name *trigonometry* for this subject even though the basic functions are called *circular* functions (Fig. 29.3). Pitiscus showed how to determine the parts into which a side of a triangle is divided by the altitude, given the lengths of the three sides, or, conversely, to determine one side of a triangle knowing the other two sides and the length of the portion of the third side cut off by the altitude. To guarantee that the angles adjacent to the side were acute, he stated the theorem only for the altitude from the vertex of the largest angle.

Pitiscus' way of deriving his fundamental relation was as follows. If the shortest side of the triangle ABC is AC and the longest is BC , let the altitude to BC be AG , as in Fig. 29.4. Draw the circle through C with center at A , so that B lies outside the circle, and let the intersections of the circle with AB and BC be E and F , respectively. Then extend BA to meet the circle at D , and connect CD . Then $\angle BFE$ is the supplement of $\angle CFE$. But $\angle EDC$ is also supplementary to $\angle CFE$, since the two are inscribed in arcs that partition the circle. Thus, $\angle BFE = \angle CDB$, and so the triangles BCD and BEF are similar. It follows that $\overline{BD} \cdot \overline{BE} = \overline{BF} \cdot \overline{BC}$, and since $\overline{BD} = \overline{AB} + \overline{AD}$, $\overline{BE} = \overline{AB} - \overline{AE}$, $\overline{AE} = \overline{AC} = \overline{AD}$, and $\overline{CF} = 2\overline{CG}$, we find

$$\overline{AB}^2 - \overline{AC}^2 = \overline{BC} \cdot \overline{BF} = \overline{BC}^2 - \overline{BC} \cdot \overline{CF} = \overline{BC}^2 - 2\overline{BC} \cdot \overline{CG}.$$

⁸This law says that the sines of the sides of spherical triangles are proportional to the sines of their opposite angles. (Both sides and angles in a spherical triangle are measured in great-circle degrees.)

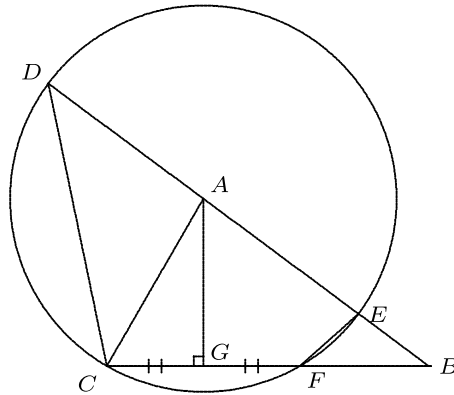


Figure 29.4. Pitiscus’ derivation of the proportions in which an altitude divides a side of a triangle.

Observe that $\overline{CG} = \overline{AC} \cos(\angle ACB)$. When this substitution is made, we obtain what is now known as the *law of cosines*:

$$\overline{AB}^2 = \overline{AC}^2 + \overline{BC}^2 - 2\overline{BC} \cdot \overline{AC} \cos(\angle C).$$

Pitiscus also gave an algebraic solution of the trisection problem discovered by an earlier mathematician named Jobst Bürgi (1552–1632). The solution had been based on the fact that the chord of triple an angle is three times the chord of the angle minus the cube of the chord of the angle. This relation makes no sense in terms of geometric dimension; it is a purely numerical relation. It is interesting that it is stated in terms of chords, since Pitiscus surely knew about sines.

29.6. A MATHEMATICAL SKILL: PROSTHAPHÆRESIS

Pitiscus needed trigonometry in order to do astronomy, especially to solve spherical triangles. Since the computations in such problems often become rather lengthy, Pitiscus discovered (probably in the writings of other mathematicians) a way to shorten the labor. While the difficulty of addition and subtraction grows at an even, linear rate with the number of digits being added, multiplying two n -digit numbers requires on the order of $2n^2$ separate binary operations on integers. Thus the labor becomes excessive and error-prone for integers with any appreciable number of digits. As astronomy becomes more precise, of course, the number of digits to which quantities can be measured increases. Thus a need arose some centuries ago for a shorter, less error-prone way of doing approximate computations.

The ultimate result of the search for such a method was the subject of logarithms. That invention, however, required a new and different point of view in algebra. Before it came along, mathematicians had found a way to make a table of sines serve the purpose that was later fulfilled by logarithms. Actually, the process could be greatly simplified by using only a table of cosines, but we shall follow Pitiscus, who used only a table of sines and thus was forced to compute the complement of an angle where we would simply look up the cosine. The principle is the same: converting a product to one or two additions and

subtractions—hence the name *prosthaphæresis*, from *prosthæresis* (taking forward, that is, addition) and *aphæresis* (taking away, that is, subtraction).

As just pointed out, the amount of labor involved in multiplying two numbers increases in direct proportion to the product of the numbers of digits in the two factors, while the labor of adding increases in proportion to the number of digits in the smaller number. Thus, multiplying two 15-digit numbers requires over 200 one-digit multiplications, and another 200 or so one-digit additions, while adding the two numbers requires only 15 such operations (not including carrying). It was the large number of digits in the table entries that caused the problem in the first place, but the key to the solution turned out to be in the structural properties of the sine function.

There are hints of this process in several sixteenth-century works, but we shall quote just one example. In his *Trigonometria*, first published in Heidelberg in 1595, Pitiscus posed the following problem: *To solve the proportion in which the first term is the radius, while the second and third terms are sines, avoiding multiplication and division.* The problem here is to find the fourth proportional x , satisfying $r : a = b : x$, where r is the radius of the circle and a and b are two sines (half-chords) in the circle. We can see immediately that $x = ab/r$, but as Pitiscus says, the idea is to avoid the multiplication and division, since in the trigonometric tables the time a and b might easily have seven or eight digits each.

The key to *prosthaphæresis* is the well-known formula

$$\sin \alpha \sin(90^\circ - \beta) = \frac{\sin(\alpha + \beta) + \sin(\alpha - \beta)}{2}.$$

This formula is applied as follows: If you have to multiply two large numbers, find two angles having the numbers as their sines. Replace one of the two angles by its complement. Next, add the angles and take the sine of their sum to obtain the first term; then subtract the angles and take the sine of their difference to obtain a second term. Finally, divide the sum of these last two sines by 2 to obtain the product. To take a very simple example, suppose that we wish to multiply 155 by 36. A table of trigonometric functions shows that $\sin(8^\circ 55') = 0.15500$ and $\sin(90^\circ - 68^\circ 54') = 0.36000$. Hence, since we moved the decimal points a total of five places to the left in the two factors, we obtain

$$36 \cdot 155 = 10^5 \frac{\sin(77^\circ 49') + \sin(-59^\circ 59')}{2} = \frac{97748 - 86588}{2} = 5580.$$

In general, some significant figures will be lost in this kind of multiplication. Obviously, no labor is saved in this simple example, but for large numbers this procedure really does make things easier. In fact, multiplying even two seven-digit numbers would tax the patience of most modern people, since it would require about 100 separate multiplications and additions. A further advantage is that *prosthaphæresis* is less error-prone than multiplication. Its advantages were known to the Danish astronomer Tycho Brahe (1546–1601),⁹ who used

⁹The formula for the product of two sines had been discovered in 1510 by Johann Werner (1468–1522). This formula and the similar formula for cosines were first published in 1588 in a small book entitled *Fundamentum astronomicum* written by Nicolai Reymers Baer (dates uncertain), known as Ursus, which is the Latin translation of Baer. Brahe, however, had already noticed their application in spherical trigonometry and had been using them during the 1580s. He even claimed credit for developing the technique himself. The origin of the technique of *prosthaphæresis* is complicated and uncertain. A discussion of it was given by Thoren (1988).

it in the astronomical computations connected with the precise observations he made at his observatory during the latter part of the sixteenth century.

This process could be simplified by using the addition and subtraction formula for cosines rather than sines. That formula is

$$\cos \alpha \cos \beta = \frac{\cos(\alpha + \beta) + \cos(\alpha - \beta)}{2}.$$

29.7. ALGEBRA: PACIOLI AND CHUQUET

The fourteenth century, in which Nicole d'Oresme made such remarkable advances in geometry and nearly created analytic geometry, was also a time of rapid advance in algebra, epitomized by Antonio de' Mazzinghi (ca. 1353–1383). His *Trattato d'algebra* (*Treatise on Algebra*) contains some complicated systems of linear and quadratic equations in as many as three unknown (Franci, 1988). He was one of the earliest algebraists to move the subject toward the numerical and away from the geometric interpretation of problems.

29.7.1. Luca Pacioli

In the fifteenth century, Luca Pacioli wrote *Summa de arithmetica, geometrica, proportioni et proportionalita* (*Encyclopedia of Arithmetic, Geometry, Proportion, and Proportionality*), which was closer to the elementary work of al-Khwarizmi and more geometrical in its approach to algebra than was the work of Mazzinghi. Actually, (Parshall, 1988) the work was largely a compilation of the works of Leonardo of Pisa, but it did bring the art of abbreviation closer to true symbolic notation. For example, what we now write as $x - \sqrt{x^2 - 36}$ was written by Pacioli as

$$1.co.\tilde{m}Rv.1.ce \tilde{m}36.$$

Here *co* means *cosa* (*thing*), the unknown. It is a translation of the Arabic word used by al-Khwarizmi. The abbreviation *ce* means *censo* (*power*), and *Rv* is probably a printed version of *Rx*, from the Latin *radix*, meaning *root*.¹⁰ Pacioli's work was both an indication of how widespread knowledge of algebra had become by this time and an important element in propagating that knowledge even more widely. The sixteenth-century Italian algebraists who moved to the forefront of the subject and advanced it far beyond where it had been up to that time had all read Pacioli's treatise thoroughly.

29.7.2. Chuquet

According to Flegg (1988), on whose work the following exposition is based, there were several new things in the *Triparty*. One is a superscript notation similar to the modern notation for the powers of the unknown in an equation. The unknown itself is called the

¹⁰The symbol *Rx* should not be confused with the same symbol in pharmacy, which comes from the Latin *recipe*, meaning *take*.

premier or “first,” that is, power 1 of the unknown. In this work, algebra is called the *rigle des premiers* “rule of firsts.” Chuquet listed the first 20 powers of 2 and pointed out that when two such numbers are multiplied, their indices are added. Thus, he had a clear idea of the laws of integer exponents. A second innovation in the *Triparty* is the free use of negative numbers as coefficients, solutions, and exponents. Still another innovation is the use of some symbolic abbreviations. For example, the square root is denoted R^2 (R for the Latin *radix*, or perhaps the French *racine*). The equation we would write as $3x^2 + 12 = 9x$ was written $.3.^2 \tilde{p}.12.$ egaulx a $.9.^1$. Chuquet called this equation impossible, since its solution would involve taking the square root of -63 .

His instructions are given in words. For example (Struik, 1986, p. 62), consider the equation

$$R^2 4.^2 \tilde{p}.4.^1 \tilde{p}.2.^1 \tilde{p}.1 \text{ egaulx a } .100,$$

which we would write

$$\sqrt{4x^2 + 4x} + 2x + 1 = 100.$$

Chuquet says to subtract $.2.^1 \tilde{p}.1$ from both sides, so that the equation becomes

$$R^2 4.^2 \tilde{p}.4.^1 \text{ egaulx a } .99\tilde{m}.2.^1.$$

Next he says to square, getting

$$4.^2 \tilde{p}.4.^1 \text{ egaulx a } 9801.\tilde{m}.396.^1 \tilde{p}.4.^2.$$

Subtracting $4.^2$ (that is, $4x^2$) from both sides and adding $396.^1$ to both sides then yields

$$400.^1 \text{ egaulx a } .9801..$$

Thus $x = 9801/400$.

Chuquet’s approach to algebra and its application can be gathered from one of the illustrative problems in the second part (Problem 35). This problem tells of a merchant who buys 15 pieces of cloth, spending a total of 160 ecus. Some of the pieces cost 11 ecus each, and the others 13 ecus. How many were bought at each price?

If x is the number bought at 11 ecus apiece, this problem leads to the equation $11x + 13(15 - x) = 160$. Since the solution is $x = 17 \frac{1}{2}$, this means the merchant bought $-2 \frac{1}{2}$ pieces at 13 ecus. How does one set about buying a negative number of pieces of cloth? Chuquet said that these $2 \frac{1}{2}$ pieces were bought on credit!

PROBLEMS AND QUESTIONS

Mathematical Problems

- 29.1.** Carry out Leonardo’s description of the way to find two numbers the sum of whose squares is a square that is the sum of two other given squares in the particular case when the given numbers are $.a. = 5$ and $.b. = 12$ (the sum of whose squares

is $169 = 13^2$). Take $.de. = 8$ and $.ez. = 15$. Draw the right triangle described by Leonardo, and also carry out the numerical computation that produces the new pair for which the sum of the squares is again 169.

- 29.2.** Use Pitiscus' law of cosines to find the third side of a triangle having sides of length 6 cm and 8 cm and such that the altitude to the side of length 8 cm divides it into lengths of 5 cm and 3 cm. (There are two possible triangles, depending on the orientation.)
- 29.3.** Use *prosthaphæresis* to find the product 829.038×66.9131 . (First write this product as $10^5 \times 0.829038 \times 0.669131$. Find the angles that have the last two numbers as cosines, and use the addition and subtraction formula for cosines given above.)

Historical Questions

- 29.4.** What parts of the algebraic work of Leonardo of Pisa were compilations of work in earlier sources, and what parts were advances on that earlier work?
- 29.5.** In what ways did the geometric work of Nicole of Oresme prefigure modern analytic geometry?
- 29.6.** How did Regiomontanus and Pitiscus change the way mathematicians thought about trigonometry? How did their trigonometry continue to differ from what we use today?

Questions for Reflection

- 29.7.** Was there scientific value in making use of the *real* (irrational, infinitely precise) number Φ , as the Bravais brothers did, even though no actual plant grows exactly according to the rule they stated? Why wouldn't a rational approximation have done just as well?
- 29.8.** How did the notion of geometric dimension (length, area, volume) limit the use of numerical methods in geometry? How did Oresme's latitude of forms help to overcome this limitation?
- 29.9.** Does it make sense to interpret the purchase of a negative number of items as an amount bought on credit? Would it be better to interpret such a "purchase" as returned merchandise?

Sixteenth-Century Algebra

Several important innovations in algebra and computation occurred during the sixteenth and early seventeenth centuries. In Italy, cubic and quartic equations were solved. In France, a new kind of notation made it possible to discuss equations in general without having to resort to specific examples, and in Scotland the discovery of logarithms reduced the labor involved in long computations.

30.1. SOLUTION OF CUBIC AND QUARTIC EQUATIONS

In Europe, algebra was confined to linear and quadratic equations for many centuries, whereas the Chinese and Japanese had not hesitated to attack equations of any degree. The difference in the two approaches is a result of different ideas of what constitutes a solution. This distinction is easy to make nowadays: The European mathematicians were seeking a sequence of arithmetic operations, including root extractions, that could be applied to the coefficients of a polynomial equation in order to produce a root, what is called *solution by radicals*, while the Chinese and Japanese were seeking the decimal expansions of real roots, one digit at a time.

The Italian algebraists of the early sixteenth century made advances in the search for a general algorithm for solving higher-degree equations. We discussed the interesting personal aspects of the solution of cubic equations in Chapter 29. Here we concentrate on the technical aspects of the solution. Because the notation of the time is rather cumbersome, on this subject we are going to use some anachronistic modern notation in order to explain the solution.

The verses Tartaglia had memorized (see Chapter 28) say, in modern language, that to solve the equation $x^3 + px = q$ for x , one should look for two numbers u and v satisfying $u - v = q$, $uv = (p/3)^3$. The problem of finding u and v is that of finding two numbers given their difference and their product; and, of course, this is merely a matter of solving a *quadratic* equation, a problem that had already been completely solved some 2500 years earlier. Once this quadratic has been solved, the solution of the original cubic is $x = \sqrt[3]{u} - \sqrt[3]{v}$. The solution of the cubic has thus been reduced to solving a quadratic equation, taking the cube roots of its two roots, and subtracting. Cardano illustrated with the case of “a cube and six times the side equal to 20.” Using his complicated rule (complicated because he

stated it in words), he gave the solution as

$$\sqrt[3]{\sqrt{108} + 10} - \sqrt[3]{\sqrt{108} - 10}.$$

He did not add that this number equals 2.

30.1.1. Ludovico Ferrari

Cardano's student Ludovico Ferrari worked with him in the solution of the cubic, and between them they had soon found a way of solving certain fourth-degree equations. Ferrari's solution of the quartic was included near the end of Cardano's treatise *Ars magna*. Counting cases as for the cubic, one finds a total of 20 possibilities. The principle in most cases is the same, however. The idea is to make a perfect square in x^2 equal to a perfect square in x by adding the same expression to both sides. Cardano gives the example

$$60x = x^4 + 6x^2 + 36.$$

It is necessary to add to both sides an expression $rx^2 + s$ to make them squares, that is, so that both sides of

$$rx^2 + 60x + s = x^4 + (6 + r)x^2 + (36 + s)$$

are perfect squares. Now the condition for this to happen is well known: $ax^2 + bx + c$ is a perfect square if and only if $b^2 - 4ac = 0$. Hence we need to have simultaneously

$$3600 - 4sr = 0, \quad (6 + r)^2 - 4(36 + s) = 0.$$

Solving the second of these equations for s in terms of r and substituting in the first leads to the equation

$$r^3 + 12r^2 = 108r + 3600.$$

This is a cubic equation called the *resolvent* cubic. Once it is solved, the original quartic breaks into two quadratic equations upon taking square roots and adding an ambiguous sign.

A few aspects of the solution of cubic and quartic equations should be noted. First, the problem is not a practical one. Second, the Cardano recipe for solving an equation sometimes gives the solution in a rather strange form. For example, Cardano says that the solution of $x^3 + 6x = 20$ is $\sqrt[3]{\sqrt{108} + 10} - \sqrt[3]{\sqrt{108} - 10}$. The expression is correct, but can you tell at a glance that it represents the number 2?

Third, the procedure does not always seem to work. For example, the equation $x^3 + 6 = 7x$ has to be solved by guessing a number that can be added to both sides so as to produce a common factor that can be canceled out. The number in this case is 21, but there is no *algorithm* for finding such a number.¹ For equations of this type, the algebraic procedure

¹There is an algorithm for finding all *rational* solutions of an equation with rational coefficients; but when the roots are irrational, this problem remains.

described by Cardano for finding x involves square roots of negative numbers. This was the first time mathematicians had encountered a need for such roots. When they occur in the solution of a quadratic equation, the roots themselves are complex numbers, making it possible to say that the equation simply has no solution. In the case of cubic equations with real coefficients, however, the algebraic procedures lead to complex numbers precisely when there are three real roots. Cardano tried to make some sense out of this case, pointing out that if one *imagined* a solution, it was possible to find solutions to quadratic equations that had previously been believed to have no roots. He gave as an example the problem of finding two numbers whose sum is 10 and whose product is 40, in other words, solving the quadratic equation $x^2 - 10x + 40 = 0$, and he gave the solutions as $5 + \sqrt{-15}$ and $5 - \sqrt{-15}$. Thus, under the influence of algebra, the stock of numbers was enlarged to include negative numbers (called *false roots* at first) and imaginary and complex numbers.

The case of three real roots came to be known as the *irreducible case* of the cubic. Strenuous efforts were made to avoid the use of complex numbers in this case, but careful analysis (see Chapter 37) showed that they are unavoidable. The difference between cubic and quadratic equations shows up in the fact that extracting the square root, and hence also the fourth root, of a complex number can be reduced to repeated extractions of the square roots of positive real numbers. But no such reduction exists for cube roots. When the equation $(x + yi)^3 = a + bi$ is written with real and imaginary parts separated, the result is generally a cubic equation for x having three distinct roots and a cubic equation for y having three distinct roots. Any attempt to remove complex numbers from the case when there are three real roots merely replaces one such equation with two others.

30.2. CONSOLIDATION

There were two natural ways to build on what had been achieved in algebra by the end of the sixteenth century. One was to find a notation that could unify equations so that it would not be necessary to consider so many different cases and so many different possible numbers of roots. The other was to solve equations of degree five and higher. We shall discuss the first of these here, reserving the second for Chapter 37.

All original algebra treatises written up to and including the treatise of Bombelli (to be discussed in Chapter 41) are very tiresome for the modern student, who is familiar with symbolic notation. For that reason we have sometimes allowed ourselves the convenience of modern notation when doing so will not distort the thought process involved too severely. In the years between 1575 and 1650, several innovations in notation were introduced that make treatises written since that time appear essentially modern. The symbols $+$ and $-$ were originally used in bookkeeping in warehouses to indicate excess and deficiencies; they first appeared in a German treatise on commercial arithmetic in 1489 but were not widely used in the rest of Europe for another century. The sign for equality was introduced by a Welsh medical doctor, physician to the short-lived Edward VI, named Robert Recorde (1510–1558). His symbol was a very long pair of parallel lines, because, as he said, “noe .2. thynges, can be moare equalle.” The use of abbreviations for the various powers of the unknown in an equation was eventually achieved, but there were two other needs to be met before algebra could become a mathematical subject on a par with geometry: a unified way of writing equations and a concept of number in which every equation would have a solution. The use of exponential notation and grouping according to powers was discussed by Simon Stevin, who was mentioned in Chapter 29. Stevin used the abbreviation M for

the first unknown in a problem, *sec* for the second, and *ter* for the third. Thus (see Zeuthen, 1903, p. 95), what we would write as the equation

$$\frac{6x^3}{y} \div 2xz^2 = \frac{3x^2}{yz^2}$$

was expressed as follows: If we divide

$$6 M \textcircled{3} D \textit{sec} \textcircled{1} \text{ by } 2 M \textcircled{1} \textit{ter} \textcircled{2},$$

we obtain

$$3 M \textcircled{2} D \textit{sec} \textcircled{1} D \textit{ter} \textcircled{2}.$$

Although notation still had far to go, from the modern point of view, at least it was no longer necessary to use a different letter to represent each power of the unknown in a problem, as Leonardo of Pisa had done in his *Liber quadratorum*.

30.2.1. François Viète

The French lawyer François Viète (1540–1603), who worked as tutor in a wealthy family and later became an advisor to Henri de Navarre (the future king Henri IV), found time to study Diophantus and to introduce his own ideas into algebra. Viète is credited with several crucial advances in the subject. In his book *Artis analyticae praxis* (*The Practice of the Analytic Art*) he begins by giving the rules for powers of binomials (in words). For example, he describes the fifth power of a binomial as “the fifth power of the first [term], plus the product of the fourth power of the first and five times the second,” Viète’s notation was slightly different from ours, but is more recognizable to us than that of Stevin. He would write the equation $A^3 + 3BA = D$, where the vowel A represented the unknown and the consonants B and D were taken as known, as follows (Zeuthen, 1903, p. 98):

$$A \textit{cubus} + B \textit{planum} \textit{ in } A^3 \textit{aequatur } D \textit{solido}.$$

As this quotation shows, Viète appears to be following the tedious route of writing everything out in words and to be adhering to the requirement that all the terms in an equation be geometrically homogeneous. In other words, the notion of quantity as a pure number, as opposed to a line or a plane or solid region, had not yet been introduced.

This introduction is followed by five books of *zetetics* (research, from the Greek word *zēteîn*, meaning *seek*). The mention of “roots” in connection with the binomial expansions was not accidental. Viète studied the relation between roots and coefficients in general equations. By using vowels to represent unknowns and consonants to represent data for a problem, Viète finally achieved what was lacking in earlier treatises: a convenient way of talking about general data without having to give specific examples. Consonants could be thought of as representing numbers that would be known in any particular application of a process, but were left unspecified for purposes of describing the process itself. We might call them parameters. His first example was the equation $A^2 + AB = Z^2$, in other words, a standard quadratic equation. According to Viète, these three letters are associated with three

numbers in direct proportion, Z being the middle, B the difference between the extremes, and A the smallest number. In our terms, $Z = Ar$ and $B = Ar^2 - A$. Thus, the general problem reduces to finding the smallest of three numbers A , Ar , Ar^2 given the middle value and the difference of the largest and smallest. Viète had already shown how to do that in his books of *zeticis*.

This analysis showed Viète the true relation between the coefficients and the roots. For example, he knew that in the equation $x^3 - 6x^2 + 11x = 6$, the sum and product of the roots must be 6 and the sum of the products taken two at a time must be 11. This observation still did not enable him to solve the general cubic equation, but he did study the problem geometrically and show that any cubic could be solved, provided that one could solve two of the classical problems of antiquity: constructing two mean proportionals between two given lines and trisecting any angle. As he concluded at the end of his geometric chapter: "It is very worthwhile to note this." In fact, by assuming the trisection of a general angle, Viète was able to avoid the annoying complex numbers that arose in the Cardano procedure for solving $x^3 + px + q = 0$ when it has three distinct real roots. In this case, the *cubic discriminant* $\frac{p^3}{27} + \frac{q^2}{4}$, whose square root needs to be taken, is negative, and that is how the complex numbers arise.

Although complex numbers began to gain acceptance after the work of Cardano and Bombelli, attempts were still made to solve the irreducible case using only real numbers. Viète's method of doing so was the most successful. It is a transcendental solution rather than an algebraic one, since it involves the cosine function.

The classical problem of trisecting the angle reduces to a cubic equation through the trigonometric identity

$$\cos^3\left(\frac{\theta}{3}\right) - \frac{3}{4}\cos\left(\frac{\theta}{3}\right) - \frac{1}{4}\cos\theta = 0.$$

This cubic equation generally has three real roots in the variable $y = \cos(\theta/3)$.

Viète's technique for solving the equation $x^3 + px + q = 0$ when there are three real roots involves "fitting" a scaled version of x to this basic equation for the cosine. Specifically, one must set $y = (\sqrt{3}/(2\sqrt{-p}))x$. (The negative sign is necessary because the existence of three distinct real roots implies that $p < 0$.) The result is the equation

$$y^3 - \frac{3}{4}y - \frac{1}{4}\frac{3\sqrt{3}q}{2p\sqrt{-p}} = 0.$$

The number $y = \cos(\theta/3)$ will be a solution of this equation if the angle θ satisfies

$$\cos\theta = \frac{3\sqrt{3}q}{2p\sqrt{-p}}.$$

There will be such an angle, provided that the right-hand side of this last equation lies between -1 and $+1$; that is, its square is at most 1. That condition amounts to $27q^2/(-4p^3) \leq 1$ and can be rewritten as $q^2/4 + p^3/27 \leq 0$. But the left-hand side of this inequality is precisely the number whose square root must be taken when following the Cardano procedure! In other words, this solution works precisely in the irreducible case. Thus, we have a non-

algebraic (transcendental) solution of the irreducible case of the cubic that does not involve any complex numbers.

30.3. LOGARITHMS

While Viète was revolutionizing algebraic notation, the problem of simplifying laborious multiplications, divisions, root extractions, and the like, was being attacked at the same time in another part of the world and from another point of view. The connection between geometric and arithmetic proportion had been noticed earlier by Chuquet, but the practical application of this fact had never been worked out. The Scottish laird John Napier, Baron of Murchiston (1550–1617), tried to clarify this connection and apply it. His work consisted of two parts: (a) a theoretical part based on a continuous geometric model and (b) a computational part involving a discrete (tabular) approximation of the continuous model. The computational part was published in 1614. However, Napier hesitated to publish his explanation of the theoretical foundation. Only in 1619, two years after his death, did his son publish the theoretical work under the title *Mirifici logarithmorum canonis descriptio* (*A Description of the Marvelous Rule of Logarithms*). The word *logarithm* means *ratio number*, and it was from the concept of ratios (geometric progressions) that Napier proceeded.

To explain his ideas, Napier used the concept of moving points. He imagined one point P moving along a straight line from a point T toward a point S with decreasing velocity such that the ratio of the distances from the point P to S at two different times depends only on the difference in the times. (Actually, he called the line ending at S a sine and imagined it shrinking from its initial size TS , which he called the radius.) A second point is imagined as moving along a second line at a constant velocity equal to that with which the first point began. These two motions are accurately drawn in Fig. 30.1.

The first point sets out from T at the same time and with the same speed with which the second point sets out from t . The first point, however, slows down, while the second point continues to move at constant speed. The figure shows the locations reached at various times by the two points: When the first point is at A , the second is at a ; when the first point is at B , the second is at b ; and so on. The point moving with decreasing velocity requires a certain amount of time to move from T to A , the same amount of time to move from A to B , from B to C , and from C to D . The geometric decrease means that $\overline{TS}/\overline{AS} = \overline{AS}/\overline{BS} = \overline{BS}/\overline{CS} = \overline{CS}/\overline{DS}$.

The first point will never reach S , since it keeps slowing down, and its velocity at S would be zero. The second point will travel indefinitely far, given enough time. Because the points are in correspondence, the division relation that exists between two positions in the first case is mirrored by a subtractive relation in the corresponding positions in the second case. Thus, this diagram essentially changes division into subtraction and multiplication

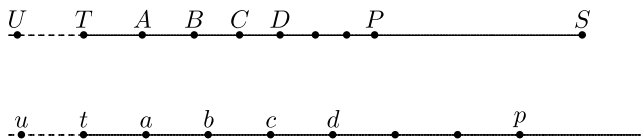


Figure 30.1. Geometric basis of logarithms.

into addition. The top scale in Fig. 30.1 resembles a slide rule, and this resemblance is not accidental: A slide rule is merely an analog computer that incorporates a table of logarithms.

Napier's definition of the logarithm can be stated in the modern notation of functions by writing $\log(\overline{AS}/\overline{TS}) = 1 = \overline{ta}/\overline{ia}$, $\log(\overline{BS}/\overline{TS}) = \overline{tb}/\overline{ia}$, and so on; in other words, the logarithm increases as the "sine" decreases. These considerations contain the essential idea of logarithms. The quantity Napier defined is not the logarithm as we know it today. If points T , A , and P correspond to points t , a , and p , then

$$\overline{tp} = \log_k \left(\frac{\overline{PS}}{\overline{TS}} \right) \cdot \overline{ia},$$

where $k = \overline{AS}/\overline{TS}$. For the computational table that he compiled, Napier took $k = 0.9999999 = 1 - 10^{-7}$.

30.3.1. Arithmetical Implementation of the Geometric Model

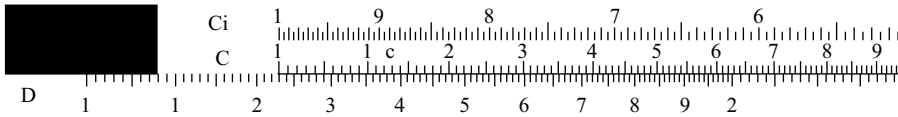
The geometric model just discussed is theoretically perfect, but of course one cannot put the points on a line into a table of numbers. It is necessary to construct the table from a finite set of points; and these points, when converted into numbers, must be rounded off. Napier analyzed the maximum errors that could arise in constructing such a table. In terms of Fig. 30.1, he showed that \overline{ia} satisfies

$$\overline{TA} < \overline{ia} < \overline{TA} \left(1 + \frac{\overline{TA}}{\overline{AS}} \right).$$

These inequalities are simple to prove. The first one is obvious, since starting from time $t = 0$, the upper point moves from T to A with velocity that is smaller than the velocity of the point below it, which is moving from t to a . As for the second, we imagine the two motions extended into the time before the lower point was at t by the same amount of time that was required for the points to reach A and a . At that earlier time, the upper point would have been at a point U , where $\overline{US}/\overline{TS} = \overline{TS}/\overline{AS}$. Consequently $\overline{UT}/\overline{TS} = \overline{US}/\overline{TS} - 1 = \overline{TS}/\overline{AS} - 1 = \overline{TA}/\overline{AS}$. Since the velocity of the upper point was larger throughout this time interval, $\overline{ia} = \overline{tu} < \overline{UT} = \overline{TS}(\overline{TA}/\overline{AS}) = \overline{TA}(1 + \overline{TA}/\overline{AS})$.

The tabular value for the logarithm of $\overline{AS}/\overline{TS}$ can be taken as the average of the two extremes, that is, as $\overline{TA}(1 + \overline{TA}/(2\overline{AS}))$, and the relative error will be very small when \overline{TA} is small.

Napier's death at the age of 67 prevented him from making some improvements in his system, which are sketched in an appendix to his treatise. These improvements consist of scaling in such a way that the logarithm of 1 is 0 and the logarithm of 10 is 1, which is the basic idea of what we now call *common logarithms*. These further improvements to the theory of logarithms were made by Henry Briggs (1561–1630), who was in contact with Napier for the last two years of Napier's life and wrote a commentary on the appendix to Napier's treatise. As a consequence, logarithms to base 10 came to be known as *Briggsian logarithms*.



Portions of the C, D, and CI scales of a slide rule. Adjacent numbers on the C and D scales are in proportion, so that $1 : 1.23 :: 1.3 : 1.599 :: 1.9 : 2.337$. Thus, the position shown here illustrates the multiplication $1.23 \cdot 1.3 = 1.599$, the division $1.722 \div 1.4 = 1.23$, and many other computations. Some visual error is inevitable. The CI (inverted) scale gives the reciprocals of the numbers on the C scale, so that division can be performed as multiplication, only using the CI scale instead of the C scale. Decimal points have to be provided by the user.

30.4. HARDWARE: SLIDE RULES AND CALCULATING MACHINES

The fact that logarithms change multiplication into addition and that addition can be performed mechanically by sliding one ruler along another led to the development of rulers with the numbers arranged in proportion to their logarithms (slide rules).

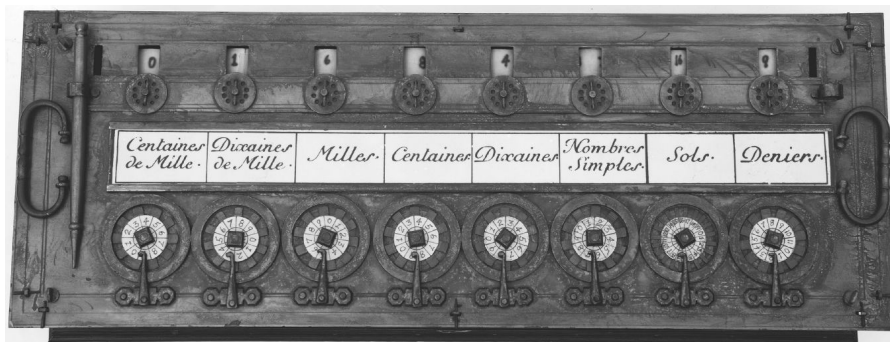
30.4.1. The Slide Rule

When one such scale is slid along a second, the numbers pair up in proportion to the distance slid, so that if 1 is opposite 5, then 3 will be opposite 15. Multiplication and division are then just as easy to do as addition and subtraction would be. The process is the same for both multiplication and division, as it was in the Egyptian graphical system, also based on proportion. Napier designed a system of rods for this purpose. The twentieth-century refinement of this idea is the slide rule.

A variant of this linear system was a system of sliding circles. Such a circular slide rule was described in a pamphlet entitled *Grammelogia* written in 1630 by Richard Delamain (1600–1644), a mathematics teacher living in London. Delamain urged the use of this device on the grounds that it made it easy to compute compound interest. Two years later the English clergyman William Oughtred (1574–1660) produced a similar description of a more complex device. Oughtred’s *circles of proportion*, as he called them, gave sines and tangents of angles in various ranges on eight different circles. Because of their portability, slide rules remained the calculating machine of choice for engineers for 350 years, and improvements were still being made in them in the 1950s. Different types of slide rule even came to have different degrees of prestige, according to the number of scales incorporated into them.

30.4.2. Calculating Machines

Slide rule calculations are floating-point computations—that is, the user has to keep track of the location of the decimal point—with limited precision and unavoidable roundoff error. When computing with integers, we often need an exact answer. To achieve that result, adding machines and other digital devices have been developed over the centuries. An early design for such a device with a series of interlocking wheels can be found in the notebooks of Leonardo da Vinci (1452–1519). Similar machines were designed by Blaise Pascal (1623–1662) and Gottfried Wilhelm Leibniz (1646–1716). Pascal’s machine was a simple adding machine that depended on turning a crank a certain number of times in order to find a sum.



A replica of Pascal's adding machine constructed by Roberto Guatelli (1904–1993). Copyright © Richard Marks. Courtesy of The Computer History Museum.

Leibniz used a variant of this machine with a removable set of wheels that would multiply, provided that the user kept count of the number of times the crank was turned.

PROBLEMS AND QUESTIONS

Mathematical Problems

- 30.1.** Verify (using a calculator) that the expression given for a root of $x^2 + 6x = 20$ really is the number 2. If you didn't have a calculator, how would you demonstrate this fact convincingly to someone?
- 30.2.** Solve the problem that Viète solved, finding all three of the numbers A , Ar , and Ar^2 , given that the middle term Ar is 15 and the difference between the largest and smallest is 40.
- 30.3.** Find $\sqrt[5]{53}$ by first finding the logarithm of 53, dividing it by 5, and taking the antilogarithm of the result. Use a calculator to do it in two ways, first with the LOG function, so that the antilogarithm of x is 10^x ; then use the LN function, so that the antilogarithm of x is e^x . Finally, check your work by computing $53^{1/5} = 53^{0.2}$ directly on the calculator.

Historical Questions

- 30.4.** Who were the main figures involved in the solution of cubic and quartic equations, and what did each of them do?
- 30.5.** What contributions to algebra are due to François Viète?
- 30.6.** In what way were logarithms an improvement on *prosthaphæresis*? Are there any situations in which one might prefer *prosthaphæresis*?

Questions for Reflection

- 30.7.** The general problem of solving a quadratic equation with complex coefficients reduces through the quadratic formula to the extraction of one square root (which may

be the square root of a complex number) followed by simple additions or subtractions and division. Extracting the square root of a complex number $a + bi$ amounts to solving simultaneously the equations $x^2 - y^2 = a$ and $2xy = b$, and these can be reduced to taking two square roots of positive real numbers. Hence there exists a purely real procedure for solving any quadratic equation when the roots are real. Taking the cube root of $a + bi$, however, means simultaneously solving $x^3 - 3xy^2 = a$, $3x^2y - y^3 = b$. In general, there are three pairs of real numbers (x, y) that will satisfy these two equations simultaneously, but finding them, as noted above, requires introducing complex numbers yet again. In what sense, then, has the general cubic equation been solved?

- 30.8.** Why was the introduction of special letters to denote constants and variables an important advance in algebra?
- 30.9.** We have seen that multiplication and division can be reduced to addition and subtraction in two different ways, namely *prosthaphæresis* and logarithms. What can you infer from this fact about the relation between trigonometric, logarithmic, and exponential functions?

Renaissance Art and Geometry

It is said that Euclid's geometry is tactile rather than visual, since the theorems tell you what you can measure and feel with your hands, not what your eye sees. It is a commonplace that a circle seen from any position except a point on the line through its center perpendicular to its plane appears to be an ellipse. If figures did not distort in this way when seen in perspective, we would have a very difficult time navigating through the world. We are so accustomed to adjusting our judgments of what we see that we usually recognize a circle automatically when we see it, even from an angle. The distortion is an essential element of our perception of depth.

31.1. THE GREEK FOUNDATIONS

Renaissance geometers and artists built on a foundation that the Greeks had laid for them in the areas of regular and semiregular solids, conic sections, and perspective. These last two have a particularly intimate relationship.

To take the simpler subject first, Euclid had discussed the five regular solids in the last three books of his *Elements*, finding the proportions between their edges and the radii of their inscribed and circumscribed spheres and showing that there can be only five such solids. He also discussed some semiregular solids such as prisms and pyramids, which would be needed later in applications of the method of exhaustion to spheres. The Renaissance geometers and artists expanded the limited circle of ideas around regular polyhedra to a large number of semiregular figures.

Euclid also wrote a book on optics in which he discussed the apparent reduction in size of objects as they move away from the observer. He used his "tactile" geometry to compare their actual sizes with their apparent sizes. The apparent sizes of the two equal vertical lines DG and AB in Fig. 31.1, as seen by the eye located at E , is measured by the angles they subtend, and thus are proportional to the arcs TH and TZ or, equivalently, to the circular sectors THE and TZE . But $DE : BE :: DZ : BA :: DZ : DG$. This last ratio, in turn, is proportional to the areas of the triangles EDZ and EDG . But EDZ is *smaller* than sector TZE , while EZG is *larger* than sector ZHE . It follows that $EZG : EDZ > ZHE : TZE$, and adding 1 to both sides implies that $EDG : EDZ > THE : TZE$. Thus, the ratio of the apparent size of DG to the apparent size of AB , which is the ratio $THE : TZE$, is *smaller*

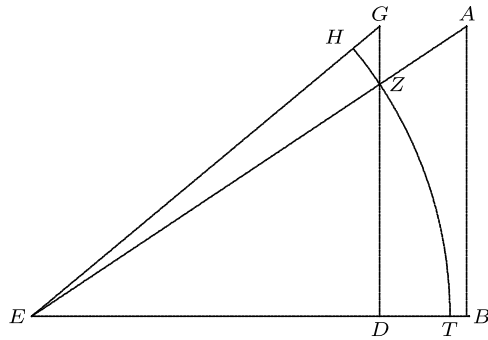


Figure 31.1. Apparent shrinkage of a vertical line as its distance from the observer increases.

than the ratio of EDG to EDZ , and hence also smaller than the ratio of BE to DE . In other words, the apparent “shrinkage” with distance is *not* inversely proportional to the distance. An object moved so as to double its distance from the eye appears to be *more than* half of its previous size.

Artists, especially those of the Italian Renaissance, used these principles to create paintings that were astoundingly realistic. As Leonardo da Vinci (1452–1519) said, “the primary task of a painter is to make a flat plane look like a body seen in relief projecting out of it.” Many records of the principles by which this effect was achieved have survived, including treatises of Leonardo himself and a painter’s manual by Albrecht Dürer (1471–1528), first published in 1525. Over a period of several centuries, these principles gave rise to the subject now known as projective geometry.

31.2. THE RENAISSANCE ARTISTS AND GEOMETERS

The revival of interest in ancient culture in general during the Renaissance naturally carried with it an interest in geometry. The famous artist Piero della Francesca (ca. 1410–1492) was inspired by the writings of Leonardo of Pisa and others to write treatises on arithmetic and the five regular solids. The scholar Luca Pacioli (1445–1517), who was influenced by Piero della Francesca and was a friend of Leonardo da Vinci, published a comprehensive treatise on arithmetic and geometry in 1494, and a second book, *De divina proportione*, in 1509. He gave the name *Divine Proportion* to what is now called the *Golden Section*, the division of a line into mean and extreme ratios. Interest in the five regular solids branched out into an interest in semiregular solids. Leonardo da Vinci designed wooden models of these, which were depicted in Pacioli’s treatise. A typical example is the truncated icosahedron, formed by cutting off the 12 vertices of an icosahedron so as to produce 12 pentagons. If the vertices are cut off at just the right distance, the middle portions of the edges of the original triangles will be exactly equal to the sides of the pentagons that replace the vertices of the triangles, so that the remaining portion of the triangular face will become a regular hexagon. The resulting figure, consisting of 20 hexagons and 12 pentagons, is the truncated icosahedron.

The regular and semiregular solids formed an important part of Dürer’s 1525 manual for painters. He showed how to cut out a paper model of a truncated icosahedron (Fig. 31.2).

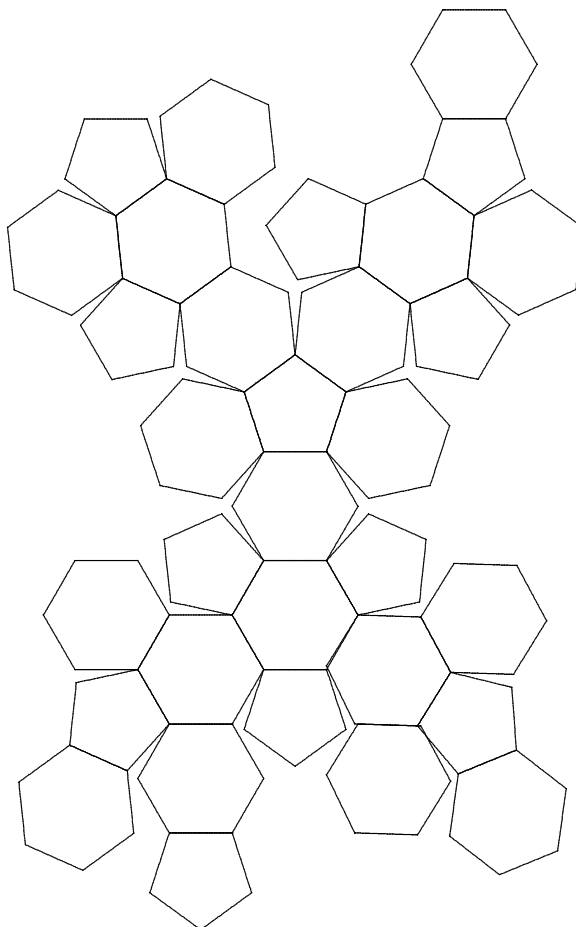
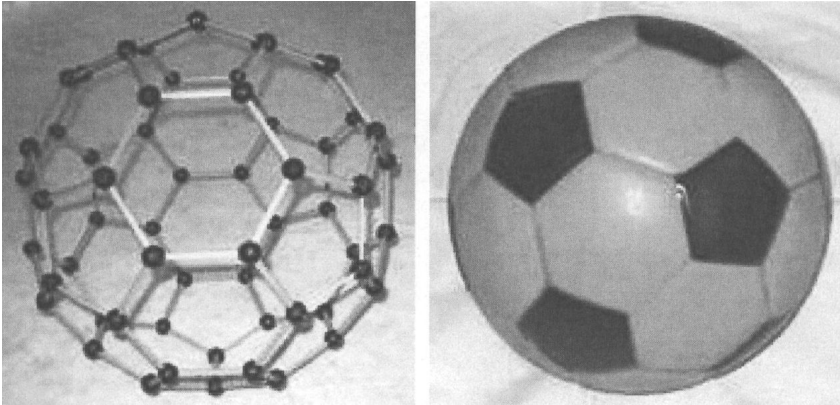


Figure 31.2. Dürer's paper model of a truncated icosahedron.

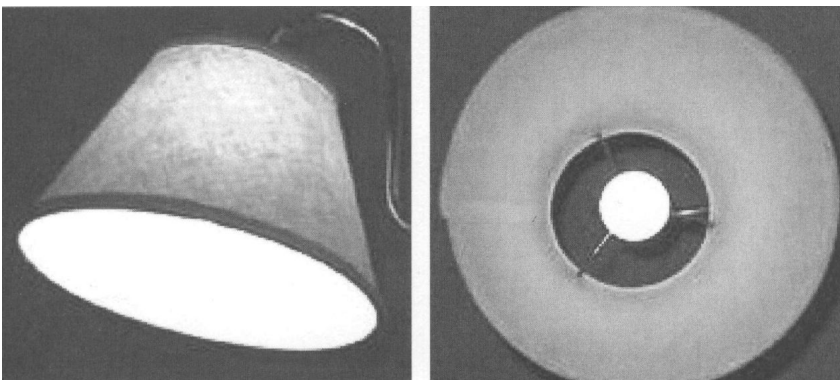
The solid, although not the name, has become very familiar to modern people through its application in athletics (the ball used in playing soccer) and organic chemistry (a molecule of C_{60} , known as buckminsterfullerene).

31.3. PROJECTIVE PROPERTIES

Projective geometry studies the relations among figures that remain constant in perspective. Among these things are points and lines, the number of intersections of lines and circles, and consequently also such things as parallelism (but not always, unless certain "points at infinity" remain at infinity) and tangency, but not things that depend on shape, such as angles or circles. Like the result from Euclid's *Optics* discussed above, its methods come out of Euclidean geometry, but explore new implications of the Euclidean definitions and assumptions.



Two modern applications of the truncated icosahedron: a molecule of C_{60} , buckminsterfullerene (a “buckyball”); a soccer ball.



A circle seen in perspective is an ellipse.

A nonobvious property that is preserved in projection is what is now called the *cross-ratio* of four points on a line.¹ If A , B , C , and D are four points on a line, with B and C both between A and D and C between B and D , their cross-ratio is

$$(A, B, C, D) = \frac{AC \cdot BD}{AD \cdot BC}.$$

If the rays PA , PB , PC , and PD from a point P intersect a second line in points A' , B' , C' , and D' , the cross-ratio of these new points is the same as that of the original four points. Coolidge (1940, p. 88) speculated that Euclid may have known about the cross-ratio, and he asserted that the early second-century mathematician Menelaus did know about it. The concept was introduced by Pappus in Book 7 of the *Synagōgē* (Jones, 1986). It is preserved

¹Although this ratio has been used for centuries, the name it now bears in English seems to go back only to an 1869 treatise on dynamics by William Kingdon Clifford (1845–1879). Before that it was called the *anharmonic ratio*, a phrase translated from an 1837 French treatise by Michel Chasles (1809–1880).

by fractional-linear transformations such as those introduced by Newton (see Chapter 38) and is used in defining distance in the models of non-Euclidean geometry (see Chapter 40).

A geometric description of perspective was written by Leon Battista Alberti (1404–1472) in 1435 in a Latin treatise entitled *De pictura*, reworked by him in Italian the following year as *Della pittura* and published posthumously in 1511. If the eye is at fixed height above a horizontal plane, parallel horizontal lines in that plane receding from the imagined point where the eye is located can be drawn as rays emanating from a point (the vanishing point) at the same height above the plane, giving the illusion that the vanishing point is infinitely distant. The application to art is obvious: Since the canvas can be thought of as a window through which the scene is viewed, if you want to draw parallel horizontal lines as they would appear through a window, you must draw them as if they all converged on the vanishing point. Thus, a family of lines having a common property (being parallel to one another) projects to a family having a different common property (passing through a common point). Obviously, lines remain lines under such a projection. However, perpendicular lines will not remain perpendicular, nor will circles remain circles.

31.3.1. Girard Desargues

The mathematical development of the theory of projection began with the work of Girard Desargues (1593–1662). In 1636, Desargues published a pamphlet with the ponderous title *An Example of One of the General Methods of S.G.D.L.² Applied to the Practice of Perspective Without the Use of Any Third Point, Whether of Distance or Any Other Kind, Lying Outside the Work Area*. The reference to a “third point” was aimed at the primary disadvantage of Alberti’s rules, the need to use a point not on the canvas in order to get the perspective correct. Three years later he produced a *Rough Draft of an Essay on the Consequences of Intersecting a Cone with a Plane*. In both works, written in French rather than the more customary Latin, he took advantage of the vernacular to invent new names, not only for the conic sections,³ as Dürer had done, but also for a large number of concepts that called attention to particular aspects of the distribution and proportions of points and lines. He was particularly fond of botanical names⁴ and included *tree*, *trunk*, *branch*, *shoot*, and *stem*, among many other neologisms. Although the new language might seem distracting, using standard terms for what he had in mind would have been misleading, since the theory he was constructing unified concepts that had been distinct before. For example, he realized that a cylinder could be regarded as a limiting case of a cone, and so he gave the name *scroll* to the class consisting of both surfaces. Desargues had very little need to refer to any specific conic section; his theorems applied to all of them equally. As he said (Field and Gray, 1987, p. 102—I have changed their *roll* to *scroll*):

The most remarkable properties of the sections of a scroll are common to all types, and the names *Ellipse*, *Parabola*, and *Hyperbola* have been given them only on account of matters extraneous to them and to their nature.

²Sieur Girard Desargues Lyonnois.

³He gave the standard names to the conic sections themselves, but suggested *deficit*, *equalation*, and *exceedence* as alternatives.

⁴Ivins (1947, cited by Field and Gray, 1987, p. 62) suggested that these names were inspired by similar names in Alberti’s treatise.

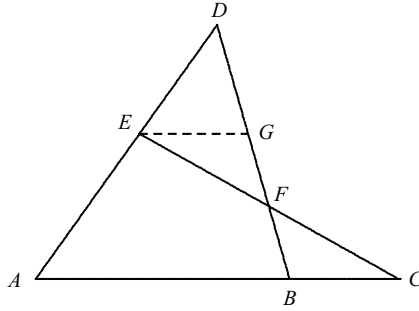


Figure 31.3. Menelaus' theorem for a plane triangle.

Desargues was among the first to regard lines as infinitely long, in the modern way. In fact, he opens his treatise by saying that he will consider both the infinitely large and the infinitely small in his work, and he says that “in this work every straight line is, if necessary, taken to be produced to infinity in both directions.” He also had the important insight that a family of parallel lines and a family of lines with a common point of intersection have similar properties. He said that lines belonged to the same *order*⁵ if either they all intersected at a common point or were all mutually parallel. This term was introduced “[to] indicate that in the one case as well as in the other, it is *as if* they all converged to the same place” [emphasis added].

Although Desargues' terminology was messy, his *Rough Draft* contained some elegant theorems about points on conics. Two significant results are the following⁶:

First: *If four lines in a plane intersect two at a time, and the points of intersection on the first line are A, B, and C, with B between A and C, and the lines through A and B intersect in the point D, those through A and C in E and those through B and C in F, then*

$$\frac{BD}{BF} = \frac{AD}{AE} \cdot \frac{CE}{CF}. \quad (31.1)$$

The situation here was described by Pappus, and the result is also known as Menelaus' theorem. The proof involves drawing the line through E parallel to AB , meeting BD in a point G , and then using the similarity of triangles EGF and CBF and of triangles DEG and DAB , as in Fig. 31.3. From Eq. (31.1) it is easy to deduce that $BD \cdot AE \cdot CF = BF \cdot AD \cdot CE$. Klein (1926, p. 80) attributes this form of the theorem to Lazare Carnot (1753–1823).

Second: *The converse of this statement is also true, and can be interpreted as stating that three points lie on a line.* That is, if ADB is a triangle, and E and F are points on AD and BD respectively such that $AD : AE < BD : BF$, then the line through E and F meets

⁵Now called a *pencil* or *sheaf*.

⁶To keep the reader's eye from getting *too* tangled up, we shall use standard letters in the statement and figure rather than Desargues' weird mixture of uppercase and lowercase letters and numbers, which almost seems to anticipate the finest principles of computer password selection.

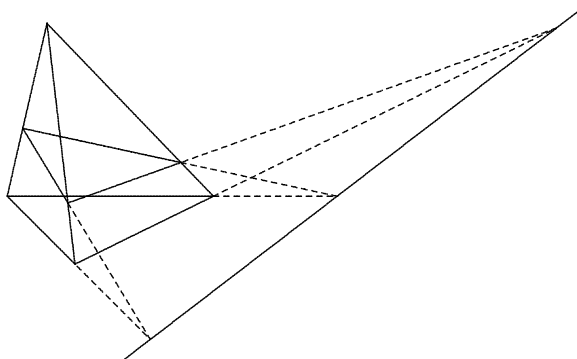


Figure 31.4. Desargues' theorem for triangles lying in different planes.

the extension of AB on the side of B in a point C , which is characterized as the only point on the line EF satisfying Eq. (31.1).

In 1648 the engraver Abraham Bosse (1602–1676), who was an enthusiastic supporter of Desargues' new ideas, published *La perspective de Mr Desargues*, in which he reworked these ideas in detail. Near the end of the book he published the theorem that is now known as Desargues' theorem. Like Desargues' work, Bosse's statement of the theorem is a tangled mess involving ten points denoted by four uppercase letters and six lowercase letters. The points lie on nine different lines. When suitably clarified, the theorem states that if the lines joining the three pairs of vertices from two different triangles intersect in a common point, the pairs of lines containing the corresponding sides of these triangles meet in three points all on the same line. This result is easy to establish if the triangles lie in different planes, since the three points must lie on the line of intersection of the two planes containing the triangles, as shown in Fig. 31.4.

For two triangles in the same plane, the theorem, illustrated in Fig. 31.5, was proved by Bosse by applying Menelaus' theorem to the three sets of collinear points $\{A'', C, B\}$, $\{B'', A, C\}$, and $\{C'', A, B\}$, with K as the third vertex of the triangle whose base ends in the second and third points in all three cases. (There is no other conceivable way to proceed, so that in a sense the proof is a mere computation.) When the ratios $AK : AA'$, $BK : BB'$, and $CK : CC'$ are eliminated from the three equations that result, what is left is the equation

$$\frac{C''B'}{C''A'} = \frac{A''B'}{A''C'} \cdot \frac{B''C'}{B''A'}.$$

Having received a copy of this work from the Parisian mathematician Marin Mersenne (1588–1648), René Descartes (1596–1654) took the word *draft* literally and regarded it as a proposal to write a treatise—which it may have been—such as a modern author would address to a publisher, and a publisher would send to an expert for review. He wrote to Desargues to express his opinion of “what I can conjecture of the *Treatise on Conic Sections*, of which [Mersenne] sent me the *Draft*.” Descartes' “review” of the work contained the kind of advice reviewers still give: that the author should decide more definitely who the intended

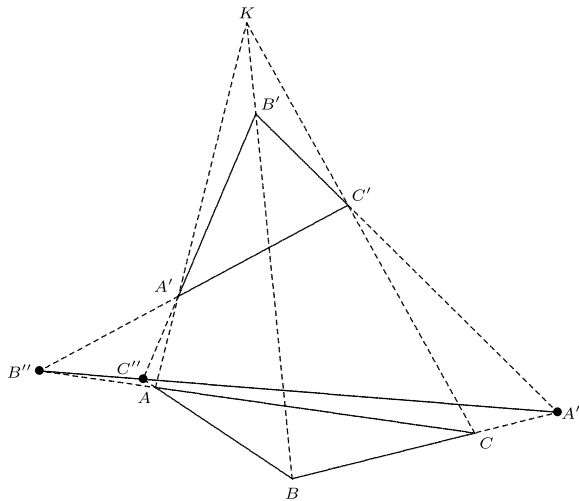


Figure 31.5. Desargues' theorem for two triangles in the same plane.

audience was. As he said, if Desargues was aiming to present new ideas to scholars, there was no need to invent new terms for familiar concepts. On the other hand, if the book was aimed at the general public, it would need to be very thick, since everything would have to be explained in great detail (Field and Gray, 1987, p. 176).

31.3.2. Blaise Pascal

Desargues' work was read by a teenage boy named Blaise Pascal (1623–1662), who was to become famous for his mathematical work and renowned for his *Pensées* (*Meditations*), which are still read by many people today for inspiration. He began working on the project of writing his own treatise on conics. Being very young, he was humble and merely sketched what he planned to do, saying that his mistrust of his own abilities inclined him to submit the proposal to experts, and “if someone thinks the subject worth pursuing, we shall try to carry it out to the extent that God gives us the strength.” Pascal admired Desargues' work very much, saying that he owed “what little I have discovered to his writings” and would imitate Desargues' methods, which he considered especially important because they treated conic sections without introducing the extraneous axial section of the cone. He used much of Desargues' notation for points and lines, including the word *order* for a family of concurrent lines. His work, like that of Desargues, remained only a draft, although Struik (1986, p. 165) reports that Pascal worked further on this project and that Leibniz saw a manuscript of it—not the rough draft, apparently—in 1676. All that has been preserved, however, is the rough draft. That draft contains several results in the spirit of Desargues, one of which, called by Pascal a “third lemma,” is well known. In the notation of Fig. 31.6, where four lines MK , MV , SK , and SV are drawn and then a conic is passed through K and V meeting these four lines in four other points P , O , N , and Q respectively, Pascal asserted that the lines PQ , NO , and MS would be concurrent (belong to the same *order*).

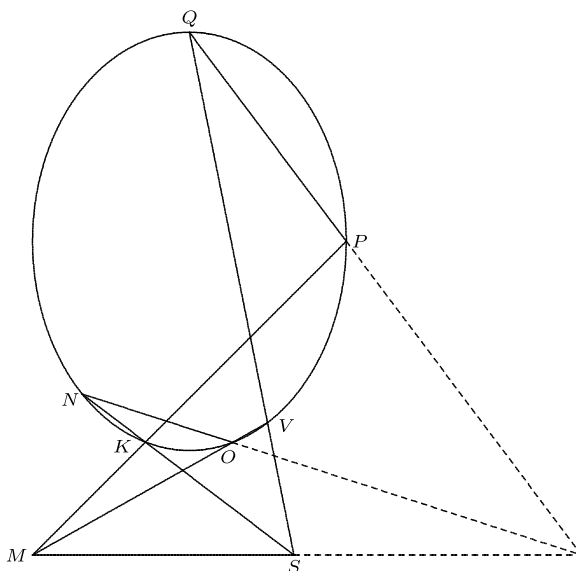


Figure 31.6. Pascal's third lemma.

PROBLEMS AND QUESTIONS

Mathematical Problems

- 31.1.** Prove Menelaus' theorem and its converse. What happens if the points E and F are such that $AD : AE :: BD : BE$? (Euclid gave the answer to this question.)
- 31.2.** Use Menelaus' theorem to prove that two medians of a triangle intersect in a point that divides each in the ratio of 1:2. You may assume that all three medians are concurrent, as this is not difficult to prove.
- 31.3.** Euclid showed that the angle subtended by an object of height h at distance d is larger at larger distances than it would be if it were merely inversely proportional to d . In other words, there is no constant k such that

$$\alpha = k \cdot \frac{h}{d},$$

and in fact $\alpha d/h$ is an increasing function of d when h is fixed.⁷ Show that, if α is measured in degrees, then

$$\lim_{d \rightarrow \infty} \frac{\alpha d}{h} = \frac{\pi}{180}.$$

⁷In other words, the actual horizontal distance grows faster than the apparent height, measured by α , shrinks. The effect is that horizontal distances appear foreshortened, and things don't appear to be as far away as they actually are. This phenomenon is well-known to anyone who has ever swum across a lake.

Historical Questions

- 31.4. Which Renaissance artists studied geometry for the purpose of creating paintings, buildings, and sculpture?
- 31.5. How does projective geometry differ from Euclidean geometry?
- 31.6. Where was the modern notion of a line as “infinitely long” first stated?

Questions for Reflection

- 31.7. What considerations may have led the Renaissance artists to renew their interest in geometry and apply it to their art?
- 31.8. One of Dürer’s devices for drawing in perspective involves a thread passing through a rectangular frame with a gate hinged on one side of it containing the canvas. On the opposite side of the frame from the object being painted, the thread passes over a pulley, which constitutes the vanishing point for the perspective. The end of the thread is held against a point to be mapped onto the canvas. Crosshairs are stuck on the frame to mark the point where the thread passes through it. The thread is then removed, the gate is closed, and a mark is made at the point where the crosshairs meet. Painting in this way is a two-person job, requiring one person to hold the end of the thread against the object and a second person to set the crosshairs. Which of these two people counts as the artist, and which as the assistant?
- 31.9. In what sense are lines and points treated symmetrically in projective geometry. (Think of Desargues’ theorem.)

The Calculus Before Newton and Leibniz

The infinite occurs in three forms in calculus: the derivative, the integral, and the power series. Integration, in the form of finding areas and volumes, was developed as a particular theory before the other two subjects came into general use. Although infinite series appear on the horizon, so to speak, in the work of Archimedes on the quadrature of the parabola, as we saw in Chapter 14, they do not come into full view.

As we have also seen, infinitesimal methods were used in geometry by the Chinese and Japanese, and the latter also used infinite series to solve geometric problems (somewhat later than Newton and Leibniz, however). In India, mathematicians had used infinite series a few centuries before the Europeans began to use them, to solve geometric problems via trigonometry. According to Rajagopal (1993), the mathematician Nilakanta, who lived in South India and whose dates are given as 1444–1543, gave a general proof of the formula for the sum of a geometric series. The most advanced of these results is attributed to Madhava (1340–1425), but is definitively stated in the work of Jyesthadeva (1530–ca. 1608):

The product of the given Sine and the radius divided by the Cosine is the first result. From the first, . . . etc., results obtain. . . a sequence of results by taking repeatedly the square of the Sine as the multiplier and the square of the Cosine as the divisor. Divide . . . in order by the odd numbers one, three, etc. . . From the sum of the odd terms, subtract the sum of the even terms. [The result] becomes the arc. [Rajagopal, 1993, p. 98]

These instructions give in words an algorithm that we would write as the following formula, remembering that the Sine and Cosine used in earlier times correspond to our $r \sin \theta$ and $r \cos \theta$, where r is the radius of the circle:

$$r\theta = \frac{r^2 \sin \theta}{r \cos \theta} - \frac{r^4 \sin^3 \theta}{3r^3 \cos^3 \theta} + \frac{r^6 \sin^5 \theta}{5r^5 \cos^5 \theta} - \cdots$$

The bulk of calculus was developed in Europe during the seventeenth century, and it is on that development that the rest of this chapter is focused.

32.1. ANALYTIC GEOMETRY

The creation of what we now know as analytic geometry had to wait for algebraic thinking about geometry (the type of thinking Pappus called *analytic*) to become a standard mode

of thinking. No small contribution to this process was the creation of the modern notational conventions, many of which were due to François Viète and René Descartes. It was Descartes who started the very useful convention of using letters near the beginning of the alphabet for constants and data and those near the end of the alphabet for variables and unknowns. Viète's convention, which was followed by Fermat, had been to use consonants and vowels respectively for these purposes.

32.1.1. Pierre de Fermat

Besides working in number theory, Fermat (1601–1665) studied the works of Apollonius, including references by to lost works. This study inspired him to write a work on plane and solid loci, first published with his collected works in 1679. He used these terms in the sense of Pappus: A plane locus is one that can be constructed using straight lines and circles, and a solid locus is one that requires conic sections for its construction. He says in the introduction that he hopes to systematize what the ancients, known to him from Book 7 of Pappus' *Synagōgē*, had left haphazard. Pappus had written that the locus to more than six lines had hardly been touched. Thus, locus problems were the context in which Fermat invented analytic geometry.

Apart from the adherence to a dimensional uniformity that Descartes (finally!) eliminated, Fermat's analytic geometry looks much like what we are now familiar with. He stated its basic principle, asserting that the lines representing two unknown magnitudes should form an angle that would usually be assumed a right angle. He began with the equation of a straight line:¹ $Z^2 - DA = BE$. This equation looks strange to us because we automatically (following Descartes) tend to look at the Z as a variable and the A and E as constants, exactly the reverse of what Fermat intended. If we make the replacements $Z \mapsto c$, $D \mapsto a$, $A \mapsto x$, $B \mapsto b$, and $E \mapsto y$, this equation becomes $c^2 - ax = by$, and now only the exponent on c looks strange, the result of Fermat's adherence to the Euclidean niceties of dimension.

Fermat illustrated the claim of Apollonius that a locus was determined by the condition that the sum of the pairwise products of lines from a variable point to given lines is given. His example was the case of two lines, where—when the two lines are mutually perpendicular—it is the familiar rectangular hyperbola that we have now seen used many times for various purposes. Fermat wrote its equation as $ae = z^2$. He showed that the graph of any quadratic equation in two variables is a conic section.

32.1.2. René Descartes

Fermat's work on analytic geometry was not published in his lifetime, and therefore was less influential than it might have been. As a result, his contemporary René Descartes (1596–1650) is remembered as the creator of analytic geometry, and we speak of “Cartesian” coordinates, even though Fermat was more explicit about their use.

Descartes is remembered not only as one of the most original and creative modern mathematicians, but also as one of the leading voices in modern philosophy and science. Both his scientific work on optics and mechanics and his geometry formed part of his philosophy. Like Plato, he formed a grand project of integrating all of human knowledge into a single system. Also like Plato, he recognized the special place of mathematics in

¹Fermat actually wrote “ Z pl. $- D$ in A æquetur B in E .” That is, “Let $Z^2 - DA$ equal BE .”

such a system. In his *Discourse on Method*, published at Leyden in 1637, he explained that logic, while it enabled a person to make correct judgments about inferences drawn through syllogisms, did not provide any actual knowledge about the world, what is usually called empirical knowledge. In what was either a deadpan piece of sarcasm or a sincere tribute to the mystic Ramon Lull (1232–1316), he said that in the art of “Lully” it enabled a person to speak fluently about matters on which he is entirely ignorant. He seems to have agreed with Plato that mathematical concepts are real objects, not mere logical relations among words, and that they are perceived directly by the mind. In his famous attempt at doubting everything, he had brought himself back from total skepticism by deducing the principle that whatever he could clearly and distinctly perceive with his mind must be correct.

As Davis and Hersh (1986) have written, the *Discourse on Method* was the fruit of a decade and a half of hard work and thinking on Descartes’ part, after a series of three vivid dreams on the night of November 10, 1619, when he was a 23-year-old soldier of fortune. The link between Descartes’ philosophy and his mathematics lies precisely in the matter of “clear and distinct perception,” for there seems to be no other area of thought in which human ideas are so clear and distinct. As Grabiner (1995, p. 84) says, when Descartes attacked, for example, a locus problem, the answer had to be “it is this curve, it has this equation, and it can be constructed in this way.” Descartes’ *Géométrie*, which contains his ideas on analytic geometry, was published as the last of three appendices to the *Discourse*.

What Descartes meant by “clear and distinct” ideas in mathematics is shown in a method of generating curves given in his *Géométrie* that appears mechanical, but can be stated in pure geometric language. A pair of lines intersecting at a fixed point Y coincide initially (Fig. 32.1). The point A remains fixed on the horizontal line. As the oblique line rotates about Y , the point B , which remains fixed on it, describes a circle. The tangent at B intersects the horizontal line at C , and the point on the oblique line directly above C is D . The line perpendicular to the oblique line at D intersects the horizontal line at E , from which a vertical line intersects the oblique line at F , and so forth in a zigzag pattern. Descartes imagined a mechanical linkage that could actually draw these curves.

Descartes regarded determinate curves of this sort, depending on one parameter, as we would say, as legitimate to use in geometry. He offered the opinion that the opposition to “mechanical” curves by ancient Greek mathematicians arose because the mechanical curves they knew about—he mentioned the spiral of Archimedes and the quadratrix—were indeterminate. In the case of the spiral of Archimedes, which is generated by a point moving

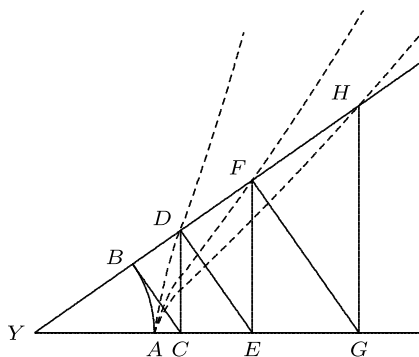


Figure 32.1. Descartes’ linkage for generating curves. The curve $x^{4n} = a^2(x^2 + y^2)^{2n-1}$ is shown for $n = 0, 1, 2, 3$.

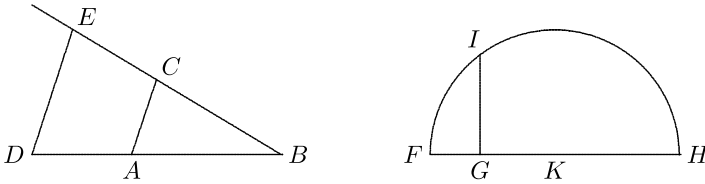


Figure 32.2. Left: $AB = 1$, so that $BE = BC \cdot BD$. Right: $FG = 1$, so that $GI = \sqrt{GH}$.

at constant linear velocity along a line that is rotating with constant angular velocity, the indeterminacy arises because the two velocities need to be coordinated with infinite precision. For the quadratrix, the same problem arises, since the ratio of the angular velocity of a rotating line and the linear velocity of a translating line needs to be known with infinite precision.

Descartes' *Géométrie* resembles a modern textbook of analytic geometry less than does Fermat's *Introduction to Plane and Solid Loci*. He does not routinely use a system of "Cartesian" coordinates, as one might expect from the name. But he does remove the dimensional difficulties that had complicated geometric arguments since Euclid's definition of a ratio.

[U]nity can always be understood, even when there are too many or too few dimensions; thus, if it be required to extract the cube root of $a^2b^2 - b$, we must consider the quantity a^2b^2 divided once by unity, and the quantity b multiplied twice by unity. [Smith and Latham, 1954, p. 6]

Here Descartes is explaining that all four arithmetic operations can be performed on *lines* and yield *lines* as a result. He illustrated the product and square root by the diagrams in Fig. 32.2, where $AB = 1$ on the left and $FG = 1$ on the right.

Descartes went a step further than Oresme in eliminating dimensional considerations, and he went a step further than Pappus in his classification of locus problems. Having translated these problems into the language of algebra, he realized that the three- and four-line locus problems always led to polynomial equations of degree at most 2 in x and y , and conversely, any equation of degree 2 or less represented a three- or four-line locus. He asserted with confidence that he had solved the problem that Pappus reported unsolved in his day. It was in this context that he formulated the idea of using two intersecting lines as a frame of reference, saying that

since so many lines are confusing, I may simplify matters by considering one of the given lines and one of those to be drawn. . . as the principal lines, to which I shall try to refer all the others. [Smith and Latham, 1954, p. 29]

The idea of using two coordinate lines is psychologically very close to the linkages illustrated in Fig. 32.1. In terms of Fig. 32.3, Descartes took one of the fixed lines as a horizontal axis AB , since a line was to be drawn from point C on the locus making a fixed angle θ with AB . He thought of this line as sliding along AB and intersecting it at point B , and he denoted the variable length AB by x . Then since C needed to slide along this moving line so as to keep the proportions demanded by the conditions of the locus problem, he denoted the distance CB by y . All the lines were fixed except CB , which moved parallel to itself, causing x to vary, while on it y adjusted to the conditions of the problem. For

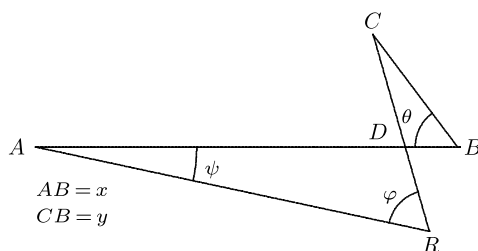


Figure 32.3. Descartes' analysis of the n -line locus problem.

each of the other fixed lines, say AR , the angles ψ , θ , and φ will all be given, ψ by the position of the fixed lines AB and AR , and the other two by the conditions prescribed in the problem. Since these three angles determine the shape of the triangles ADR and BCD , they determine the ratios of any pair of sides in these triangles through the law of sines, and hence all sides can be expressed in terms of constants and the two lengths x and y . If the set of $2n$ lines is divided into two sets of n as the $2n$ -line locus problem requires, the conditions of the problem can be stated as an equation of the form

$$p(x, y) = q(x, y),$$

where p and q are of degree at most n in each variable. The analysis was mostly “clear and distinct.”

Descartes argued that the locus could be considered known if one could locate as many points on it as desired.² He next pointed out that in order to locate points on the locus one could assign values to either variable x and y , then compute the value of the other by solving the equation.³

Everyone who has studied analytic geometry in school must have been struck at the beginning by how much clearer and easier it was to use than the synthetic geometry of Euclid. That aspect of the subject is nicely captured in the words the poet Paul Valéry (1871–1945) applied to Descartes' philosophical method in general: “the most brilliant victory ever achieved by a man whose genius was applied to reducing the need for genius” (quoted by Davis and Hersh, 1986, p. 7).

This point was not appreciated by Newton, who, in a rather ungenerous exhibition of his own remarkable mathematical talent (Whiteside, 1967, Vol. IV, pp. 275–283), said that Descartes “makes a great show” about his solution of the three- and four-line locus problems, “as if he had achieved something so earnestly sought after by the ancients.” He also expressed a distaste for Descartes' use of symbolic algebra to solve this problem (a distaste that would be echoed by other mathematicians), saying that if this algebra were written out in words, it “would prove to be so tedious and entangled as to provoke nausea.” One is inclined to say, on Descartes' behalf, “Precisely! That's why it's better to use algebraic symbolism and avoid the tedium, confusion, and nausea.”

²The validity of this claim is somewhat less than “clear and distinct.”

³This claim also involves a great deal of hope, since equations of degree higher than 4 were unknown territory in his day.

32.2. COMPONENTS OF THE CALCULUS

In his comprehensive history of the calculus, Boyer (1949) described “a century of anticipation” during which the application of algebra to geometric problems began to incorporate some of the less systematic parts of ancient geometry, especially the infinitesimal ideas contained in what was called the method of indivisibles. Let us take up the story of calculus at the point where algebra enters the picture, beginning with some elementary problems of finding tangents and areas.

32.2.1. Tangent and Maximum Problems

The main problem in finding a tangent to a curve at a given point is to find some second condition, in addition to passing through the point, that this line must satisfy so as to determine it uniquely. It suffices to know either a second point that it must pass through or the angle that it must make with a given line.

Fermat had attacked the problem of finding maxima and minima of variables even before the publication of Descartes’ *Géométrie*. As his works were not published during his lifetime but only circulated among those who were in a rather select group of correspondents, his work in this area was not recognized for some time. His method is very close to what is still taught in calculus books. The difference is that whereas we now use the derivative to find the *slope* of the tangent line, that is, the tangent of the angle it makes with a reference axis, Fermat looked for the *point* where the tangent intercepted that axis. If the two lines did not intersect, as happens at maxima and minima, the tangent was easily determined as the unique parallel through the given point to the given axis. In all other cases Fermat needed to determine the length of the projection of the tangent on the axis from the point of intersection to the point below the point of tangency, a length known as the *subtangent*. In a letter sent to the monk and mathematician Marin Mersenne (1588–1648) and forwarded to Descartes in 1638, Fermat explained his method of finding the subtangent, which invokes some of the same ideas used earlier by Sharaf al-Tusi (see Chapter 26).

In Fig. 32.4 the curve DB is a parabola with axis CD , and the tangent at B , above the point C , meets the axis at E_C . Since the parabola is convex, a point O between B and E_C on the tangent lies outside the parabola. That location provided Fermat with two inequalities, one of which was $\overline{CD} : \overline{DI} > \overline{BC}^2 : \overline{OI}^2$. (Equality would hold here if \overline{OI} were replaced by the portion of it cut off by the parabola.) Since $\overline{BC} : \overline{OI} = \overline{CE_C} : \overline{E_CI}$, it follows that $\overline{CD} : \overline{DI} > \overline{CE_C}^2 : \overline{E_CI}^2$. Then abbreviating by setting $\overline{CD} = g$, $\overline{CE_C} = x$, and $\overline{CI} = y$, we have $g : g - y > x^2 : x^2 + y^2 - 2xy$. Cross-multiplying, we obtain

$$gx^2 + gy^2 - 2gxy > gx^2 - x^2y.$$

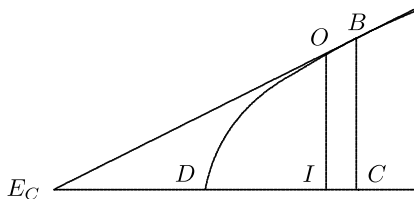


Figure 32.4. Fermat’s method of finding the subtangent.

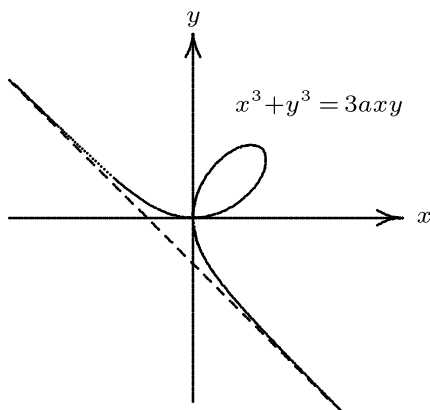


Figure 32.5. The folium of Descartes. Descartes and Fermat considered only the loop in this curve.

Canceling the term gx^2 and dividing by y , we obtain $gy - 2gx > -x^2$. Since this inequality must hold for all positive numbers y (no matter how small), it follows that $x^2 \geq 2gx$, that is, $x \geq 2g$ if $x > 0$. That is, $\overline{CE}_C \geq 2\overline{CD}$ if C is right of the point I , and so $\overline{IE}_I \geq 2\overline{CD}$ also. Reasoning similarly if C is left of I , shows that $\overline{CE}_C \leq 2\overline{CD}$ if C is any point left of I , and so $\overline{IE}_I \leq 2\overline{CD}$. Having thus ruled out the possibilities $\overline{IE}_I < 2\overline{CD}$ and $\overline{IE}_I > 2\overline{CD}$, Fermat had proved that $\overline{IE}_I = 2\overline{ID}$, and thereby solved the problem. In this argument, Fermat was relying on the Archimedean trichotomy. His argument can be formulated as a limiting argument, but it is perfectly rigorous in the finite terms in which he stated it.

In this paper Fermat asserted, “And this method never fails. . . .” This assertion provoked an objection from Descartes,⁴ who used circles in a very similar way to find the normal to a curve at a point. (We do not have space to discuss that method.) Descartes challenged Fermat with the curve of Fig. 32.5, now known as the folium of Descartes, having equation $x^3 + y^3 = 3axy$.

As already mentioned, Descartes did not regard curves such as the spiral and the quadratrix as admissible in argument, since they are generated by two motions whose relationship to each other cannot be determined exactly. A few such curves, however, were to prove a very fruitful source of new constructions and applications. One of them, which had first been noticed in the early sixteenth century by an obscure mathematician named Charles Bouvelles (ca. 1470–ca. 1553), is the cycloid, the curve generated by a point on a circle (called the generating circle) that rolls without slipping along a straight line. It is easily pictured by imagining a painted spot on the rim of a wheel as the wheel rolls along the ground. Since the speed of the rim about its center is exactly equal to the linear speed of the center, it follows that the point is at any instant moving along the bisector of the angle formed by a horizontal line and the tangent to the generating circle. In this way, given the generating circle, it is an easy matter to construct the tangent to the cycloid. This result was obtained independently around 1638 by Descartes, Fermat, and Gilles Personne de

⁴There was little love lost between Descartes and Fermat, since Fermat had dismissed Descartes’ derivation of the law of refraction. [Descartes assumed that light traveled faster in denser media; Fermat assumed that it traveled slower. Yet they both arrived at the same law! For details, see Indorato and Nastasi 1989.] Descartes longed for revenge, and even though he eventually ended the controversy over Fermat’s methods with a grudging half-acknowledgment that Fermat was right, he continued to attack Fermat’s construction of the tangent to a cycloid.

Roberval (1602–1675), and slightly later by Evangelista Torricelli (1608–1647), a pupil of Galileo Galilei (1564–1642).

32.2.2. Lengths, Areas, and Volumes

Seventeenth-century mathematicians had inherited two conceptually different ways of applying infinitesimal ideas to find areas and volumes. One was to regard an area as a “sum of lines.” The other was to approximate the area by a sum of regular figures and try to show that the approximation got better as the individual regular figures got smaller. The rigorous version of the latter argument, the method of exhaustion, based on the Archimedean trichotomy was tedious and of limited application.

32.2.3. Bonaventura Cavalieri

In the “sum of lines” approach, a figure whose area or volume was required was sliced into parallel sections, and these sections were shown to be equal or proportional to corresponding sections of a second figure whose area or volume was known. The first figure was then asserted to be equal or proportional to the second. The principle was stated in 1635 by Bonaventura Cavalieri (1598–1647), a Jesuit priest and a student of Galileo. At the time it was customary for professors to prove their worthiness for a chair of mathematics by a learned dissertation. Cavalieri proved certain figures equal by pairing off congruent sections of them, in a manner similar to Archimedes’ *Method* and the method by which Zu Chongzhi and Zu Geng found the volume of a sphere. This method implied that figures in a plane lying between two parallel lines and such that all sections parallel to those lines have the same length must have equal area. This principle is now called *Cavalieri’s principle*. The idea of regarding a two-dimensional figure as a sum of lines or a three-dimensional figure as a sum of plane figures was extended by Cavalieri to consideration of the squares on the lines in a plane figure, then to the cubes on the lines in a figure, and so on.

To see how his reasoning works, we give a sample. Figure 32.6 shows an isosceles right triangle ABC of side $AB = a = BC$, completed to form the square $ABCF$. The “sum of the lines” such as DE inside this triangle is simply its area, namely $a^2/2$. The sum of the lines in the triangle AHK , whose sides are only half as large, is one-fourth of this area,

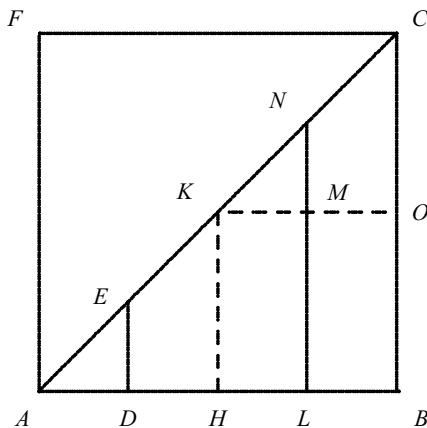


Figure 32.6. The sum of the powers of the lines in a triangle, according to Cavalieri.

or $a^2/8$. (One can see this by letting each line BE correspond to the line only half as far from A as B is. That line will be only half as long as BE , and the total length along which the new lines are “summed” will also be only half as long.) Cavalieri undertook to find a similar expression for the sum of the squares of the lines, which we would interpret as the area under the curve $y = x^2$. The sum of the portion of these squares that lie between A and H is one-eighth of the full sum, since again one can imagine summing the squares of the lines only half as far away from A , which are only one-fourth as long, and the length along which they are summed is only half of side AB .

The sum of all the squares inside ABC is the sum of all the squares DE^2 inside triangle AHK , plus the sum of all the squares LN^2 inside the trapezoid $HBCK$. But since $LN = LM + MN = a/2 + MN$, we see that the latter sum is the sum of all the squares MN^2 inside KOC (which is the same as the sum of all the squares DE^2 inside AHK) plus a times the sum of all the lines MN inside KOC (which is $a^3/8$), plus another $a^3/8$ for the sum of the squares $a^2/4$ of the lines inside the square $HBOK$. Altogether then, the sum of the squares of the lines inside ABC is twice the sum of the squares of the lines inside AHK , plus $a^3/4$. Since the sum of the squares of the lines inside AHK is one-eighth of the sum of the squares of the lines inside ABC , it follows that three fourths of the latter sum is $a^3/4$, and therefore the sum is $a^3/3$. In this way, Cavalieri established what is equivalent to the formula $\int_0^a x^2 dx = a^3/3$. More generally, using the binomial expansion of $(a/2 + MN)^n$ just as we did here in the case $n = 2$, he established the equivalent of $\int_0^a x^n dx = a^{n+1}/n$.

32.2.4. Gilles Personne de Roberval

Cavalieri’s principle was applied to find the area of the cycloid. Roberval (1602–1675), who found the tangent to the cycloid, also found the area beneath it by a clever use of Cavalieri’s principle. He considered along with half an arch of the cycloid itself a curve he called the *companion* to the cycloid. This companion curve is generated by a point that is always directly below or above the center of the generating circle as it rolls along and at the same height as the point on the rim that is generating the cycloid. As the circle makes half a revolution (see Fig. 32.7), the cycloid and its companion first diverge from the ground level, then meet again at the top. Symmetry considerations show that the area under the companion curve is exactly one-half of the rectangle whose vertical sides are the initial and final positions of the diameter of the generating circle through the point generating the cycloid. But by definition of the two curves their generating points are always at the same height, and the horizontal distance between them at any instant is half of the corresponding

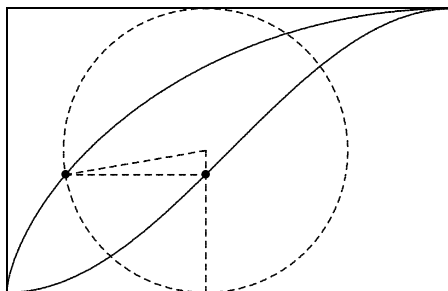


Figure 32.7. Roberval’s quadrature of the cycloid.

horizontal section of the generating circle. Hence by Cavalieri’s principle the area between the two curves is exactly half the area of the circle.

32.2.5. Rectangular Approximations and the Method of Exhaustion

Besides the method of indivisibles (Cavalieri’s principle), mathematicians of the time also applied the method of polygonal approximation, that is, the method of exhaustion to find areas. This method, as we saw in Chapter 14, was used by Archimedes to find the area of a parabolic segment and was completely rigorous from the Euclidean point of view. In 1640, Fermat wrote a paper on quadratures in which he found the areas under certain figures by this method. He was somewhat sketchier in the details than Archimedes had been, but that was because he referred explicitly to Archimedes’ work, saying that “it suffices to make this remark once and for all, and there is no need to refer constantly to a technique that is well known to mathematicians.” In other words, as we see from examining what Fermat wrote, all the basic ideas of the analysis were already present in the work of Archimedes, and what had been lacking was an algebraic language to make the expression of those ideas more transparent. That language is what Fermat supplied.

He considered a “generalized hyperbola,” as in Fig. 32.8, a curve referred to asymptotes AR and AC and defined by the property that the ratio $AH^m : AG^m = EG^n : HI^n$ is the same for any two points E and I on the curve; we would describe this property by saying that $x^m y^n = \text{const}$.

The case $n = 1, m = 2$ is illustrated in Fig. 32.6, where the abscissas $AG, AH, AO, AM, AR, \dots$ increase geometrically, that is $AG/AH = AH/AO = AO/AM = AM/AR, \dots$, and the ordinates $EG, IH, NO, PM, SR, \dots$ are inversely proportional to the squares of the corresponding abscissas. As we would write this relation, $EG = k/AG^2$, and so on. From these relations, it is not difficult to see that the area S under the curve from G to infinity satisfies

$$\begin{aligned} S &> \overline{GH} \cdot \overline{HI} + \overline{HO} \cdot \overline{OK} + \overline{OM} \cdot \overline{MN} + \overline{MP} \cdot \overline{PR} + \dots \\ &= \frac{k \cdot \overline{GH}}{\overline{AH}^2} \left(1 + \frac{\overline{AG}}{\overline{AH}} + \left(\frac{\overline{AG}}{\overline{AH}} \right)^2 + \left(\frac{\overline{AG}}{\overline{AH}} \right)^3 + \dots \right) \\ &= \frac{k}{\overline{AH}}. \end{aligned}$$

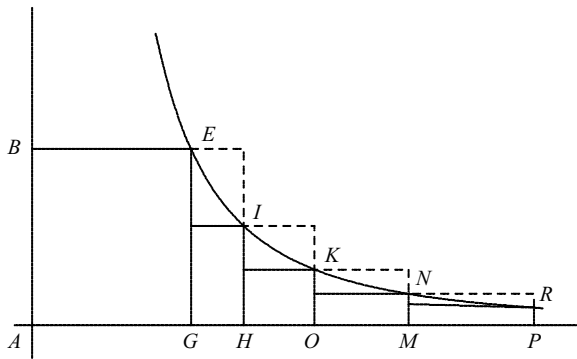


Figure 32.8. Fermat’s quadrature of a generalized hyperbola.

And similarly,

$$\begin{aligned} S &< \overline{GH} \cdot \overline{EG} + \overline{HO} \cdot \overline{HI} + \overline{OM} \cdot \overline{OK} + \overline{MP} \cdot \overline{MN} + \dots \\ &= \frac{k \cdot \overline{GH}}{\overline{AG}^2} \left(1 + \frac{\overline{AG}}{\overline{AH}} + \left(\frac{\overline{AG}}{\overline{AH}} \right)^2 + \left(\frac{\overline{AG}}{\overline{AH}} \right)^3 + \dots \right) \\ &= \frac{k \cdot \overline{AH}}{\overline{AG}^2}. \end{aligned}$$

When H is taken sufficiently close to G , both of these expressions can be made as close to $\frac{k}{\overline{AG}}$, that is, to $\overline{GE} \cdot \overline{AG}$, as desired. Hence, the rectangle $AGEB$ must be the required area. We would now phrase this result as

$$\int_a^\infty \frac{k}{x^2} dx = \frac{k}{a}.$$

32.2.6. Blaise Pascal

As shown above, Cavalieri found the “sums of the powers of the lines” inside a triangle, which makes it possible to find the area under a curve $y = x^n$ between any two values $x = a$ and $x = b$. Naturally, one would like to be able to do the same for the portion of a semicircle $h = \sqrt{R^2 - x^2}$ cut off between two chords $x = a$ and $x = b$. Thus instead of summing the lines x^n from $x = a$ to $x = b$, we would like to sum the lines $\sqrt{R^2 - x^2}$. The technique we illustrated above will not help, since the portion of the sum in the first half of the interval has no simple relation to the whole sum, and the binomial expansion of $\sqrt{R^2 - ((a/2 + MN))^2}$ in terms of MN is infinite and very messy. How is this problem to be solved? To explain it, we shall introduce a bit of later notation due to Leibniz to clarify what Cavalieri was actually doing. He wasn’t actually summing all the x^2 between $x = 0$ and $x = a$. Rather, he was summing $x^2 dx$, where dx is an “infinitely short” portion of the x -axis. The technical basis for his results, as shown in the example given above, was the use of the binomial theorem to break the sum of powers of lines inside a rectangle into the portions above and below the diagonal. That technique will not work for a circle, where instead of powers of x , one needs to consider lines of length $\sqrt{R^2 - x^2}$. To find the quadrature of the circle, it is necessary to change to a new variable, namely the polar angle φ , and express dx in terms of $d\varphi$. That feat was achieved by Blaise Pascal (1623–1662) in 1659.

Pascal found the “sums of the powers of the lines inside a quadrant of a circle.” Now a line inside a quadrant of a circle is what up to now has been called a sine. Thus, Pascal found the sum of the powers of the sines of a quadrant of a circle. Ordinarily, as we have repeatedly seen, the geometric interpretation of this sum would be as the area under the curve. However, that will not work in this case. To keep the reasoning clear (or, rather, as clear as it could be at the time, since he was using actual infinitesimal arguments, in contrast to Fermat), Pascal distinguished between a sine and an ordinate. We can keep this distinction clear by noting that it is all a matter of which variable is regarded as the independent variable. The ordinate is $y = \sqrt{R^2 - x^2}$ and refers to rectangular coordinates (x, y) . The sine is $R \sin \varphi$ and refers to what we call polar coordinates (φ, r) . Thus, the “sum of the sines” is carried out only over an infinitesimal arc, which must be imagined as having been straightened out

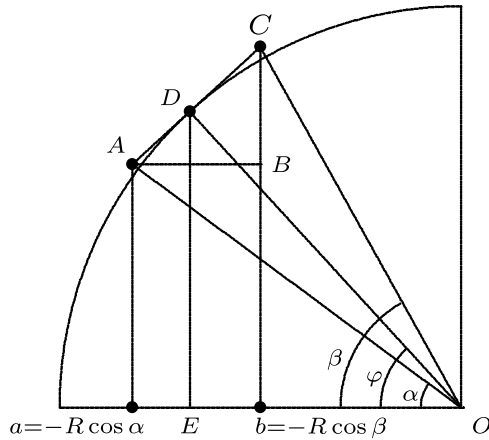


Figure 32.9. Pascal’s infinitesimal triangle method of finding the sum of the sines in a quadrant.

(rectified) and turned into a portion of the tangent.⁵ The area under the curve is the sum of the ordinates, which we can’t find. Pascal managed to express the sum of the ordinates in terms of the sum of the sines, which he was able to find.

Pascal argued that when a finite arc is divided into *infinitely many* equal small pieces, those pieces are equal to the portions of the tangent cut off between the two bounding ordinates. He acknowledged that the arc and the tangent were not equal with only a finite number of divisions, but claimed that they were equal with an infinite number. Thus he took the plunge into an actual infinity. This was a bold step beyond what the Greeks and Fermat would have admitted. This argument was logically shaky at the time, but the intuitive riches that Leibniz reaped from it ultimately justified a few centuries of doubt and uncertainty as to its validity. To see what Pascal did, refer to Fig. 32.9.

We would like to sum the infinitely thin rectangles $\sqrt{R^2 - x^2} dx$, but the algebraic techniques used by Cavalieri and Fermat do not allow us to do so. By approximating a small piece of arc of length $d(R\varphi) = R(\beta - \alpha)$ by the tangent AC , and drawing the radius AD to the point of tangency, we get an infinitesimal triangle ABC that is similar to the finite triangle DEO , whence it follows that $OD \cdot AB = AC \cdot DE$. Now, $AB = R(\cos \alpha - \cos \beta) = d(R \cos \varphi)$, $OD = R$, $AC = d(R\varphi)$, and $DE = R \sin \varphi$. Thus we have the infinitesimal relation

$$R d(R \cos \varphi) = R \sin \varphi d(R\varphi),$$

and it is clear that R can be divided out of this relation. Pascal expressed this result by saying that *the sum of the sines of any arc is equal to the portion of the base between the extreme sines [that is, $R \cos \alpha - R \cos \beta$] multiplied by the radius.*

In terms of the rectangular coordinates x and y , the differential relation says $dx = R \sin \varphi d\varphi$. Hence to find the area (the sum of the ordinates, in Pascal’s language) from $\varphi = 0$ to $\varphi = \beta$, we would have to sum $R \sin \varphi dx = R^2 \sin^2 \varphi d\varphi$. Pascal did so, and

⁵Leibniz, who was inspired by Pascal’s infinitesimal triangle, showed that the chord would work just as well as the tangent.

determined correctly that *the sum of the squares of those sines is equal to the sum of the ordinates that lie between the extreme sines.*

However, we are getting ahead of the story by talking about differential relations. As noted, the concept of a differential was due to Leibniz, inspired by precisely the work of Pascal that we are discussing here.

In modern terms, where Cavalieri found $\int_0^a x^n dx = a^{n+1}/(n+1)$, Pascal found $\int_\alpha^\beta (R \sin \varphi) R d\varphi = R(R \cos \alpha - R \cos \beta)$.

32.2.7. The Relation Between Tangents and Areas

The first statement of a relation between tangents and areas appears in 1670 in a book entitled *Lectiones geometricae* by Isaac Barrow (1630–1677), a professor of mathematics at Cambridge and later chaplain to Charles II. Barrow gave the credit for this theorem to “that most learned man, Gregory of Aberdeen” (James Gregory, 1638–1675). Barrow states several theorems resembling the fundamental theorem of calculus. The first theorem (Section 11 of Lecture 10) is the easiest to understand. Given a curve referred to an axis, Barrow constructs a second curve such that the ordinate at each point is proportional to the area under the original curve up to that point. We would express this relation as $F(x) = (1/R) \int_a^x f(t) dt$, where $y = f(x)$ is the first curve, $y = F(x)$ is the second, and $1/R$ is the constant of proportionality. If the point $T = (t, 0)$ is chosen on the axis so that $(x - t) \cdot f(x) = RF(x)$, then, said Barrow, T is the foot of the subtangent to the curve $y = F(x)$; that is, $x - t$ is the length of the subtangent. In modern language the length of the subtangent to the curve $y = F(x)$ is $|F(x)/F'(x)|$. This expression would replace $(x - t)$ in the equation given by Barrow. If both $F(x)$ and $F'(x)$ are positive, this relation really does say that $f(x) = RF'(x) = (d/dx) \int_a^x f(t) dt$.

Later, in Section 19 of Lecture 11, Barrow shows the other version of the fundamental theorem, that is, that if a curve is chosen so that the ratio of its ordinate to its subtangent (this ratio is precisely what we now call the derivative) is proportional to the ordinate of a second curve, the area under the second curve is proportional to the ordinate of the first.

32.2.8. Infinite Series and Products

The methods of integration requiring the summing of infinitesimal rectangles or all the lines inside a plane figure led naturally to the consideration of infinite series. Several special series were known by the mid-seventeenth century. For example, the Scottish mathematician James Gregory published a work on geometry in 1668 in which he stated the equivalent of the formula given earlier (unknown to Gregory, of course) by Jyesthadeva:

$$\arctan x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \dots$$

Infinite product expansions were known by this time for the number π . One, due to John Wallis (1616–1703), is

$$\frac{2}{\pi} = \frac{1 \cdot 3 \cdot 3 \cdot 5 \cdot 5 \cdot 7 \cdot \dots}{2 \cdot 2 \cdot 4 \cdot 4 \cdot 6 \cdot 6 \cdot \dots}$$

32.2.9. The Binomial Series

It was the binomial series that really established the use of infinite series in analysis. The expansion of a power of a binomial leads to finite series when the exponent is a nonnegative integer and to an infinite series otherwise. This series, which we now write in the form

$$(1 + x)^r = 1 + \sum_{k=1}^{\infty} \frac{r(r-1)\cdots(r-k+1)}{1\cdots k} x^k,$$

was discovered by Newton around 1665, although he expressed it in a different language, as a recursive procedure for finding the terms. In a 1676 letter to Henry Oldenburg (1615–1677), Secretary of the Royal Society, Newton wrote this expansion as

$$\sqrt[m]{P + PQ} = P \left| \frac{m}{n} = P \left| \frac{m}{n} + \frac{m}{n} A Q + \frac{m-n}{2n} B Q + \frac{m-2n}{3n} C Q + \frac{m-3n}{4n} D Q + \text{etc.} \right. \right.$$

“where $P + PQ$ stands for a quantity whose root or power or whose root of a power is to be found, P being the first term of that quantity, Q being the remaining terms divided by the first term and m/n the numerical index of the powers of $P + PQ$. . . A stands for the first term $P \left| \frac{m}{n} \right.$, B for the second term $\frac{m}{n} A Q$, and so on. . . .”

Newton’s explanation of the meaning of the terms A, B, C, \dots , means that the k th term is obtained from its predecessor via multiplication by $\left\{ \left[\frac{m}{n} - k \right] / (k + 1) \right\} Q$. He said that m/n could be any fraction, positive or negative.

The entrance of Newton into this arena was to lead to revolutionary changes in the way people thought about the myriad techniques and principles that made up the subject that was soon to become the calculus. That story forms the subject of our next chapter.

PROBLEMS AND QUESTIONS

Mathematical Problems

32.1. Show that the Madhava–Jyesthadeva formula given at the beginning of the chapter is equivalent to

$$\theta = \sum_{k=0}^{\infty} (-1)^k \frac{\tan^{2k+1} \theta}{2k + 1},$$

or, letting $x = \tan \theta$,

$$\arctan x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{2k + 1}.$$

32.2. Referring to Fig. 32.6, show that the rectangle containing half of the cycloidal arch has length equal to the circumference of the circle and height equal to its diameter, and so, because of Archimedes’ result that a circle equals a right triangle having

one leg equal to its circumference and the other equal to its radius, that rectangle is exactly twice as large as the circle generating the cycloid. Use that and Roberval's result to show that the area under one full arch of a cycloid is exactly three times the area of the generating circle.

- 32.3.** Consider an ellipse with semiaxes a and b ($a > b$) and a circle of radius b , the center of the circle lying on the extension of the major axis of the ellipse. Show that for every line parallel to major axis of the ellipse, the portion of that line inside the ellipse will be a/b times the portion inside the circle. Use this fact and Cavalieri's principle to compute the area of the ellipse. This result was given by Kepler.

Historical Questions

- 32.4.** What differences are noticeable between analytic geometry as developed by Fermat and by Descartes?
- 32.5.** What major innovation in the application of geometry to algebra is due to Descartes?
- 32.6.** Which infinite series were the first to be introduced, and what was the motive for introducing them?

Questions for Reflection

- 32.7.** What methodological techniques inherited from Euclidean geometry had to be ignored in order to apply algebra to geometry?
- 32.8.** The philosopher Bertrand Russell wrote (1945, p. 36), "When Descartes introduced coordinate geometry, thereby again making arithmetic supreme [above geometry], he assumed the possibility of a solution to the problem of incommensurables, though in his day no such solution had been found." What is the "problem of incommensurables" that Russell is referring to, and on what grounds can one conclude that Descartes was ignoring the difficulty?
- 32.9.** Given the large number of tangent and area problems solved by the mid-seventeenth century, what was still needed before one could say that calculus had arisen?

Newton and Leibniz

The discoveries described in the preceding chapter show that the essential components of calculus were recognized by the mid-seventeenth century, like the pieces of a jigsaw puzzle lying loose on a table. What was needed was someone to see the pattern and fit all the pieces together. The unifying principle was the concept of a derivative, and that concept came to Newton and Leibniz independently and in slightly differing forms.

33.1. ISAAC NEWTON

Isaac Newton was born prematurely on Christmas day in 1642. (When the British adopted the Gregorian calendar in 1752, eleven days had to be removed as an adjustment from the earlier Julian calendar. As a result, on what is called the *proleptic* Gregorian calendar, his actual birthday was January 4, 1643.) His parents were minor gentry, but his father had died before his birth. The midwife who attended his mother is said to have predicted that the child would not live out the day. Medical predictions are notoriously unreliable, and this one was wrong by 85 years! He was 6 years old when the English Civil War began, and the rest of his childhood was spent in that turbulent period. He attended a neighborhood school, and though not a particularly good student, he exhibited enough talent to motivate his uncle to send him to Cambridge University, which he entered about the time of the restoration of Charles II to the throne. Although he was primarily interested in chemistry, he did buy and read not only Euclid's *Elements* but also some of the current treatises on algebra and analytic geometry. From 1663 on he attended the lectures of Isaac Barrow.

Due to an outbreak of plague in 1665, he returned to his family home at Woolsthorpe, and during the next two years, while the University was closed, he alternated between Woolsthorpe and his rooms in Cambridge, pursuing his own mathematical and physical researches. He was a careful observer and experimenter, and this period was, as he later recalled, the most productive of his life. Besides the binomial theorem already discussed, he discovered the general use of infinite series and what he called the method of fluxions. His early notes on the subject were not published until after his death, but a revised version of the method was expounded in his *Principia*.

33.1.1. Newton's First Version of the Calculus

Newton first developed the calculus in what we would call parametric form. Time was the universal independent variable, and the relative rates of change of other variables were

computed as the ratios of their rates of change with respect to time. Newton thought of variables as moving quantities and focused attention on their velocities. He used the letter o to represent a small time interval and p for the velocity of the variable x , so that the change in x over the time interval o was op . Similarly, using q for the velocity of y , if y and x are related by $y^n = x^m$, then $(y + oq)^n = (x + op)^m$. Both sides can be expanded by the binomial theorem. Then if the equal terms y^n and x^m are subtracted, all the remaining terms are divisible by o . When o is divided out, one side is $nqy^{n-1} + oA$ and the other is $mpx^{m-1} + oB$. Ignoring the terms containing o , since o is small, one finds that the relative rate of change of the two variables, q/p is given by $q/p = (mx^{m-1})/(ny^{n-1})$; and since $y = x^{m/n}$, it follows that $q/p = (m/n)x^{(m/n)-1}$. Here at last was the concept of a derivative, expressed as a relative rate of change.

Newton recognized that reversing the process of finding the relative rate of change provides a solution of the area problem. He was able to find the area under the curve $y = ax^{m/n}$ by working backward.

33.1.2. Fluxions and Fluents

Newton's "second draft" of the calculus was the concept of fluents and fluxions. A *fluent* is a moving or flowing quantity; its *fluxion* is its rate of flow, which we now call its velocity or derivative. In his *Fluxions*, written in Latin in 1671 and published in 1742 (an English translation appeared in 1736), he replaced the notation p for velocity by \dot{x} , a notation still used in mechanics and in the calculus of variations. Newton's notation for the opposite operation, finding a fluent from the fluxion, is no longer used. Instead of $\int x(t) dt$, he wrote \dot{x} .

The first problem in the *Fluxions* is: *The relation of the flowing quantities to one another being given, to determine the relation of their fluxions.* The rule given for solving this problem is to arrange the equation that expresses the given relation (assumed algebraic) in increasing integer powers of one of the variables, say x , multiply its terms by any arithmetic progression (that is, the first power is multiplied by c , the square by $2c$, the cube by $3c$, etc.), and then multiply by \dot{x}/x . After this operation has been performed for each of the variables, the sum of all the resulting terms is set equal to zero.

Newton illustrated this operation with the relation $x^3 - ax^2 + axy - y^2 = 0$, for which the corresponding fluxion relation is $3x^2\dot{x} - 2ax\dot{x} + a\dot{x}y + ax\dot{y} - 2y\dot{y} = 0$, and by numerous examples of finding tangents to well-known curves such as the spiral and the cycloid. Newton also found their curvatures and areas. The combination of these techniques with infinite series was important, since fluents often could not be found in finite terms. For example, Newton found that the area under the curve $z = 1/(1 + x^2)$ was given by the Jyesthadeva–Gregory series $z = x - \frac{1}{3}x^3 + \frac{1}{5}x^5 - \frac{1}{7}x^7 + \dots$.

33.1.3. Later Exposition of the Calculus

Newton made an attempt to explain fluxions in terms that would be more acceptable logically, calling it the "method of initial and final ratios," in his treatise on mechanics, the *Philosophiae naturalis principia mathematica* (*Mathematical Principles of Natural Philosophy*), where he said the following:

Quantities, and the ratios of quantities, which in any finite time converge continually toward equality, and before the end of that time approach nearer to each other than by any given difference, become ultimately equal.

If you deny it, suppose them to be ultimately unequal, and let D be their ultimate difference. Therefore they cannot approach nearer to equality than by that given difference D ; which is contrary to the supposition.

If only the phrase *become ultimately equal* had some clear meaning, as Newton seemed to assume, this argument might have been convincing. As it is, it comes close to being a definition of *ultimately equal*, or, as we would say, equal in the limit. Newton came close to stating the modern concept of a limit at another point in his treatise, when he described the “final ratios” (derivatives) as “limits towards which the ratios of quantities decreasing without limits do always converge, and to which they approach nearer than by any given difference.” Here one can almost see the “arbitrarily small ε ” that plays the central role in the modern definition of a limit.

33.1.4. Objections

Newton anticipated some objections to these principles, and in his *Principia*, tried to phrase his exposition of the method of initial and final ratios in such a way as not to outrage anyone’s logical scruples. He said:

It may be objected that there is no final ratio of vanishing quantities, because before they vanish their ratio is not the final one, and after they vanish, they have no ratio. But that same argument would imply that a body arriving at a certain place and stopping there has no final velocity, because the velocity before it arrived was not its final velocity; and after it arrived, it had no velocity. But the answer is easy: the final velocity is the velocity the body has at the exact instant when it arrives, not before or after.

Was this explanation adequate? Do human beings in fact have any conception of what is meant by an instant of time? Do we have a clear idea of the velocity of a body *at the very instant* when it stops moving? Or do some people only imagine that we do? We are here very close to the arrow paradox of Zeno. At any given instant, the arrow does not move; therefore it is at rest. How can there be a motion (a traversal of a positive distance) as a result of an accumulation of states of rest, in each of which no distance is traveled? Newton’s “by the same argument” practically invited the further objection that his attempted explanation merely stated the same fallacy in a new way.

33.2. GOTTFRIED WILHELM VON LEIBNIZ

The codiscoverer with Newton of the calculus was, like Newton, a man involved in public life, but a much more amiable character. The philosopher Bertrand Russell, who had studied Leibniz and understood him better than anyone else, proclaimed him not an admirable man. According to Russell, Leibniz developed a profound philosophy, which he kept secret, knowing that it would not be popular, and published instead only a fatuous optimism aimed at winning friends. Leibniz, the optimistic philosopher, was parodied in the character of Dr. Paingloss in Voltaire’s *Candide*.

As was the case with Newton, Leibniz had wide-ranging interests as a youth and focused on mathematics only in early adulthood. He was born in Leipzig in 1646, more than three years later than Newton, and entered the university there in 1661, at the age of 15. Like Descartes, Fermat and Viète, he studied the law, but was considered too young to be awarded the degree of doctor of laws when he finished his course at the age of 20. He entered the service of the Elector of Mainz as a diplomat and finally came to serve the Electors of Hannover for four decades, including the future King George I of Britain, who succeeded Queen Anne in 1714. In contrast to the prickly, anti-social Newton, he was an urbane, tolerant man, who worked diplomatically in an attempt to reunite the Catholic and Protestant churches, and it was his suggestion to Tsar Peter I (1682–1726) that the Russian Academy of Sciences be founded. This was done the year before Peter died, and many talented mathematicians, including Daniel Bernoulli and Leonhard Euler, did some of their best work there.

During his lifetime, France was militarily powerful while Germany was divided and weak. As servant of several German princes, Leibniz attempted to shield Germany from the power of the French by diverting the interests of Louis XIV toward a holy war against the Ottoman Empire in Egypt. It was during a mission to Paris in 1672 that Leibniz became interested in mathematics and began to read the writings of Pascal. The following year he visited London and met some members of the Royal Society, including the secretary Henry Oldenburg and the librarian James Collins (1625–1683). He kept a diary of this journey on a sheet of paper ruled into columns headed Chemistry, Mechanica, Magnetica, Botanica, and so on. Under mathematics the notes are very sparse, containing only a reference to a general method of finding tangents, probably derived from the lectures of Barrow, which he had bought.

From this time on, Leibniz studied mathematics in earnest and within a decade had derived most of the calculus in essentially the form we know it today. His approach to the subject, in particular the delicate notion of the meaning to be assigned to the limiting ratio of two quantities as they vanish, is quite different from Newton's.

33.2.1. Leibniz' Presentation of the Calculus

Leibniz believed in the reality of infinitesimals, quantities so small that any finite sum of them is still less than any assignable positive number, but which are nevertheless not zero, so that one is allowed to divide by them. The three kinds of numbers (finite, infinite, and infinitesimal) could, in Leibniz' view, be multiplied by one another, and the result of multiplying an infinite number by an infinitesimal might be any one of the three kinds. This position was rejected in the nineteenth century but was resurrected in the twentieth century and made logically sound. It lies at the heart of what is called *nonstandard analysis*, a subject that has not penetrated the undergraduate curriculum. The radical step that must be taken in order to believe in infinitesimals is a rejection of the Archimedean principle that for any two positive quantities of the same kind, some finite number of bisections of the first will produce a quantity smaller than the second. This principle was essential to the use of the method of exhaustion, which was one of the crowning glories of Euclidean geometry. It is no wonder that mathematicians were reluctant to give it up.

Leibniz invented the expression dx to indicate the difference of two infinitely close values of x , dy to indicate the difference of two infinitely close values of y , and dy/dx to indicate the ratio of these two values. This notation was beautifully intuitive and is still the preferred notation for thinking about calculus. Its logical basis at the time was questionable, since it

avoided the objections listed above by claiming that the two quantities have not vanished at all but have yet become less than any assigned positive number. However, at the time, consistency would have been counterproductive in mathematics and science.

The integral calculus and the fundamental theorem of calculus flowed very naturally from Leibniz' approach. Leibniz could argue that the ordinates to the points on a curve represent infinitesimal rectangles of height y and width dx , and hence finding the area under the curve—"summing all the lines in the figure"—amounted to summing infinitesimal increments of area dA , which accumulated to give the total area. Since it was obvious that, on the infinitesimal level, $dA = y dx$, the fundamental theorem of calculus was an immediate consequence. Leibniz first set it out in geometric form in a paper on quadratures in the 1693 *Acta eruditorum*, a scholarly journal founded by the philosopher Otto Mencke (1644–1707) in Leipzig in 1682. In that paper, Leibniz considered two curves: one, which we would now write as $y = f(x)$, with its graph above a horizontal axis, the other, which we write as $z = F(x)$, with its graph below the horizontal axis.¹ The second curve has an ordinate proportional to the area under the first curve. That is, for a positive constant a , having the dimension of length, $aF(x)$ is the area under the curve $y = f(x)$ from the origin up to the point with abscissa x . We would write the relation now as²

$$aF(x) = \int_0^x f(t) dt.$$

In this form the relation is dimensionally consistent. What Leibniz proved was that the curve $z = F(x)$, which he called the *quadratrix* (squarer), could be constructed from its infinitesimal elements. In Fig. 33.1, the parentheses around letters denote points at an infinitesimal distance from the points denoted by the same letters without parentheses. In the infinitesimal triangle $CE(C)$ the line $E(C)$ represents dF , while the infinitesimal quadrilateral $HF(F)(H)$ represents dA , the element of area under the curve. The lines $F(F)$ and CE both represent dx . Leibniz argued that by construction, $a dF = f(x) dx$, and so $dF : dx = f(x) : a$. That meant that the quadratrix could be constructed by antidifferentiating $f(x)$.

Leibniz eventually abbreviated the sum of all the increments in the area (that is, the total area) using an elongated S, so that $A = \int dA = \int y dx$. Nearly all the basic rules of calculus for finding the derivatives of the elementary functions and the derivatives of products, quotients, and so on, were contained in Leibniz' 1684 paper on his method of finding tangents. He had obtained these results several years earlier. His collected works contain a paper written in Latin with the title *Compendium quadraturæ arithmeticae*, to which the editor assigns a date of 1678 or 1679. This paper shows Leibniz' approach through infinitesimal differences and their sums and suggests that it was primarily the problem of squaring the circle and other conic sections that inspired this work, which consists of 49 propositions and two problems. Most of the propositions are stated without proof.

¹The vertical axis is to be assumed positive in both directions from the origin. We are preserving in Fig. 33.1 only the lines needed to explain Leibniz' argument. He himself merely labeled points on the two curves with letters and referred to those letters.

²The limits of integration shown here were unknown in Leibniz' time. They were introduced in the nineteenth century by Joseph Fourier (1768–1830).

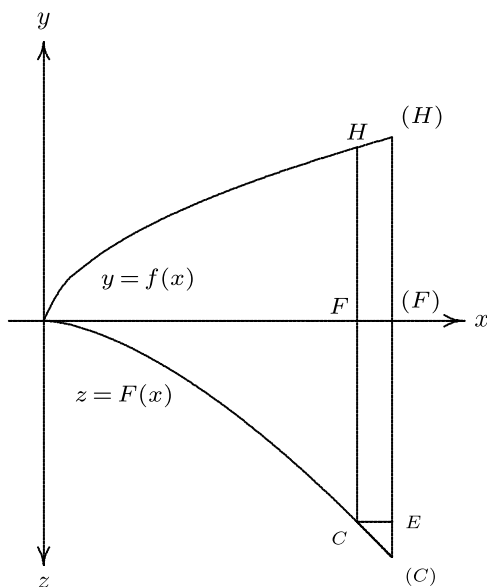


Figure 33.1. Leibniz' proof of the fundamental theorem of calculus.

Among them are the Taylor series expansions of logarithms, exponentials, and trigonometric functions.

33.2.2. Later Reflections on the Calculus

Like Newton, Leibniz felt the need to answer objections to the new methods of the calculus. In the *Acta eruditorum* of 1695 Leibniz published a "Response to certain objections raised by Herr Bernardo Nieuwentiit³ regarding differential or infinitesimal methods." These objections were three: (1) that certain infinitely small quantities were discarded as if they were zero⁴; (2) the method could not be applied when the exponent is a variable; and (3) the higher-order differentials were inconsistent with Leibniz' claim that only geometry could provide the necessary foundation. In answer to the first objection, Leibniz attempted to explain different orders of infinitesimals, pointing out that one could neglect all but the lowest orders in a given equation. To answer the second, he used the binomial theorem to demonstrate how to handle the differentials dx , dy , dz when $y^x = z$. To answer the third, Leibniz said that one should not think of $d(dx)$ as a quantity that fails to yield a (finite) quantity even when multiplied by an infinite number. He pointed out that if x varies geometrically when y varies arithmetically—in modern terms, if $x = e^{y/a}$ —then $dx = (x dy)/a$ and $ddx = (dx dy)/a$, which makes perfectly good sense.

³Bernard Nieuwentijt (1654–1718) was a Dutch Calvinist theologian.

⁴This principle was set forth as fundamental in the following year in the textbook of calculus by the Marquis de l'Hospital (1661–1704).

33.3. THE DISCIPLES OF NEWTON AND LEIBNIZ

Newton and Leibniz had disciples who carried on their work. Among Newton's followers was Roger Cotes (1682–1716), who oversaw the publication of a later edition of Newton's *Principia* and defended Newton's inverse square law of gravitation in a preface to that work. He also fleshed out the calculus with some particular results on plane loci and considered the extension of functions defined by power series to complex values, deriving the important formula $i\phi = \log(\cos \phi + i \sin \phi)$, where $i = \sqrt{-1}$. Another of Newton's followers was Brook Taylor (1685–1731), who developed a calculus of finite differences that mirrors in many ways the "continuous" calculus of Newton and Leibniz and is of both theoretical and practical use today. Taylor is famous for the infinite power series representation of functions that now bears his name. It appeared in his 1715 treatise on finite differences. We have already seen that many particular "Taylor series" were known to Newton and Leibniz; Taylor's merit is to have recognized a general way of producing such a series in terms of the derivatives of the generating function. This step, however, was also taken by Leibniz' disciple John Bernoulli.

Leibniz also had a group of active and intelligent followers who continued to develop his ideas. The most prominent of these were the Bernoulli brothers James (1654–1705) and John (1667–1748), citizens of Switzerland, between whom relations were not always cordial. They investigated problems that arose in connection with calculus and helped to systematize, extend, and popularize the subject. In addition, they pioneered new mathematical subjects such as the calculus of variations, differential equations, and the mathematical theory of probability. A French nobleman, the Marquis de l'Hospital, took lessons from John Bernoulli and paid him a salary in return for the right to Bernoulli's mathematical discoveries. As a result, Bernoulli's discovery of a way of assigning values to what are now called indeterminate forms appeared in L'Hospital's 1696 textbook *Analyse des infiniment petits* (*Infinitesimal Analysis*) and has ever since been known as L'Hospital's rule. Like the followers of Newton, who had to answer the objections of Bishop Berkeley that will be discussed in the next section, Leibniz' followers encountered objections from Michel Rolle (1652–1719), objections that were answered by John Bernoulli with the claim that Rolle didn't understand the subject.

33.4. PHILOSOPHICAL ISSUES

Some objections to the calculus were eloquently stated seven years after Newton's death by the philosopher George Berkeley⁵ (1685–1753, Anglican Bishop of Cloyne, Ireland), for whom the city of Berkeley⁶ in California is named. In his 1734 book *The Analyst*, Berkeley first took on Newton's fluxions, noting that "It is said that the minutest errors are not to be neglected in mathematics."⁷ Berkeley continues:

[It is said] that the fluxions are celerities [speeds], not proportional to the finite increments, though ever so small; but only to the moments or nascent increments, whereof the proportion alone, and not the magnitude, is considered. And of the aforesaid fluxions there be other

⁵Pronounced "Barkley."

⁶Pronounced "Birkley."

⁷It was indeed said, and by Newton himself, in his 1704 treatise *Introduction to the Quadrature of Curves*.

fluxions, which fluxions of fluxions are called second fluxions. And the fluxions of the second fluxions are called third fluxions: and so on, fourth, fifth, sixth, &c. *ad infinitum*. Now, as our sense is strained and puzzled with the perception of objects extremely minute, even so the imagination, which faculty derives from sense, is very much strained and puzzled to frame clear ideas of the least particles of time. . . and much more so to comprehend. . . those increments of the flowing quantities. . . in their very first origin, or beginning to exist, before they become finite particles. . . The incipient celerity of an incipient celerity, the nascent augment of a nascent augment, *i.e.*, of a thing which hath no magnitude: take it in what light you please, the clear conception of it will, if I mistake not, be found impossible.

He then proceeded to attack the views of Leibniz:

The foreign mathematicians are supposed by some, even of our own, to proceed in a manner less accurate, perhaps, and geometrical, yet more intelligible. . . Now to conceive a quantity infinitely small, that is, infinitely less than any sensible or imaginable quantity or than any the least finite magnitude is, I confess, above my capacity. But to conceive a part of such infinitely small quantity that shall be still infinitely less than it, and consequently though multiplied infinitely shall never equal the minutest finite quantity, is, I suspect, an infinite difficulty to any man whatsoever.

Berkeley analyzed a curve whose area up to x was x^3 (he wrote xxx). If $z - x$ was the increment of the abscissa and $z^3 - x^3$ the increment of area, the quotient would be $z^2 + zx + x^2$. He said that, if $z = x$, of course this last expression is $3x^2$, and that must be the ordinate of the curve in question. That is, its equation must be $y = 3x^2$. But, he pointed out,

[H]erein is a direct fallacy: for, in the first place, it is supposed that the abscisses z and x are unequal, without which supposition no one step could have been made [that is, the division by $z - x$ would have been undefined]; which is a manifest inconsistency, and amounts to the same thing that hath been before considered. . . The great author of the method of fluxions felt this difficulty, and therefore he gave in to those nice abstractions and geometrical metaphysics without which he saw nothing could be done on the received principles. . . It must, indeed, be acknowledged that he used fluxions, like the scaffold of a building, as things to be laid aside or got rid of as soon as finite lines were found proportional to them. . . And what are these fluxions? The velocities of evanescent increments? And what are these same evanescent increments? They are neither finite quantities, nor quantities infinitely small, nor yet nothing. May we not call them the ghosts of departed quantities?

33.4.1. The Debate on the Continent

Calculus disturbed the metaphysical assumptions of philosophers and mathematicians on the Continent as well as in Britain. L'Hospital's textbook had made two explicit assumptions: first, that if a quantity is increased or diminished by a quantity that is infinitesimal in comparison with itself, it may be regarded as remaining unchanged. Second, that a curve may be regarded as an infinite succession of straight lines. L'Hospital's justification for these claims was not commensurate with the strength of the assumptions. He merely said:

[T]hey seem so obvious to me that I do not believe they could leave any doubt in the mind of attentive readers. And I could even have proved them easily after the manner of the Ancients,

if I had not resolved to treat only briefly things that are already known, concentrating on those that are new. [Quoted by Mancosu, 1989, p. 228]

The idea that $x + dx = x$, implicit in l'Hospital's first assumption, leads algebraically to the equation $dx = 0$ if equations are to retain their previous meaning. Rolle raised this objection and was answered by the claim that dx represents the distance traveled in an instant of time by an object moving with finite velocity. This debate was carried on in private in the Paris Academy during the first decade of the eighteenth century, and members were at first instructed not to discuss it in public, as if it were a criminal case! Rolle's criticism could be answered, but it was *not* answered at the time. According to Mancosu (1989), the matter was settled in a most unacademic manner, by making l'Hospital into an icon after his death in 1704. His eulogy by Bernard Lebouyer de Fontenelle (1657–1757) simply declared the anti-infinitesimalists wrong, as if the Academy could decide metaphysical questions by fiat, just as it can define what is proper usage in French:

[T]hose who knew nothing of the mysteries of this new infinitesimal geometry were shocked to hear that there are infinities of infinities, and some infinities larger or smaller than others; for they saw only the top of the building without knowing its foundation. [Quoted by Mancosu (1989, 241)]

33.5. THE PRIORITY DISPUTE

One of the better known and less edifying incidents in the history of mathematics is the dispute between the disciples of Newton and those of Leibniz over the credit for the invention of the calculus. Although Newton had discovered the calculus by the early 1670s and had described it in a paper sent to James Collins, the librarian of the Royal Society, he did not publish his discoveries until 1687. Leibniz made his discoveries a few years later than Newton but published some of them earlier, in 1684. Newton's vanity was wounded in 1695 when he learned that Leibniz was regarded on the Continent as the discoverer of the calculus, even though Leibniz himself made no claim to this honor. In 1699 a Swiss immigrant to England, Nicolas Fatio de Duillier (1664–1753), suggested that Leibniz had seen Newton's paper when he had visited London and talked with Collins in 1673. (Collins died in 1683, before his testimony in the matter was needed.) This unfortunate rumor poisoned relations between Newton and Leibniz and their followers.

In 1711–1712 a committee of the Royal Society (of which Newton was President) investigated the matter and reported that it believed Leibniz had seen certain documents that in fact he had not seen. Relations between British and Continental mathematicians reached such a low ebb that Newton deleted certain laudatory references to Leibniz from the third edition of his *Principia*. This dispute confirmed the British in the use of the clumsy Newtonian notation for more than a century, a notation far inferior to the elegant and intuitive symbolism of Leibniz. But in the early nineteenth century the impressive advances made by Continental scholars such as Euler, Lagrange, and Laplace won over the British mathematicians. Scholars such as William Wallace (1768–1843) rewrote the theory of fluxions in terms of the theory of limits. Wallace asserted that there was never any need to introduce motion and velocity into this theory, except as illustrations, and that indeed Newton himself used motion only for illustration, recasting his arguments in terms of limits when rigor was needed [see Panteki (1987) and Craik 1999)]. Eventually, even the British began using the

term *integral* instead of *fluent* and *derivative* instead of *fluxion*, and these Newtonian terms became mathematically part of a dead language.

Some important facts were obscured by the terms in which the priority dispute was cast. One of these is the extent to which Fermat, Descartes, Cavalieri, Pascal, Roberval, and others had developed the techniques in isolated cases that were to be unified by the calculus as we know it now. In any case, Newton's teacher Isaac Barrow had the insight into the connection between subtangents and area before either Newton or Leibniz thought of it. Barrow's contributions were ignored in the heat of the dispute; their significance has been pointed out by Feingold (1993).

33.6. EARLY TEXTBOOKS ON CALCULUS

The secure place of calculus in the mathematical curriculum was established by the publication of a number of textbooks. One of the earliest was the *Analyse des infiniment petits*, mentioned above, which was published by the Marquis de l'Hospital in 1696.

Most students of calculus know the Maclaurin series as a special case of the Taylor series. Its discoverer was a Scottish contemporary of Taylor, Colin Maclaurin (1698–1746), whose *Treatise of Fluxions* (1742) contained a thorough and rigorous exposition of calculus. It was written partly as a response to Berkeley's attacks on the foundations of calculus.

The Italian textbook *Istituzioni analitiche ad uso della gioventù italiana* (*Analytic Principles for the Use of Italian Youth*) became a standard treatise on analytic geometry and calculus and was translated into English in 1801. Its author was Maria Gaetana Agnesi (1718–1799), one of the first women to achieve prominence in mathematics.

The definitive textbooks of calculus were written by the greatest mathematician of the eighteenth century, the Swiss scholar Leonhard Euler. In his 1748 *Introductio in analysin infinitorum*, a two-volume work, Euler gave a thorough discussion of analytic geometry in two and three dimensions, infinite series (including the use of complex variables in such series), and the foundations of a systematic theory of algebraic functions. The modern presentation of trigonometry was established in this work. The *Introductio* was followed in 1755 by *Institutiones calculi differentialis* and a three-volume *Institutiones calculi integralis* (1768–1774), which included the entire theory of calculus and the elements of differential equations, richly illustrated with challenging examples. It was in Euler's textbooks that many prominent nineteenth-century mathematicians such as the Norwegian genius Niels Henrik Abel (1802–1829) first encountered higher mathematics, and the influence of Euler's methods and results can be traced in their work.

33.6.1. The State of the Calculus Around 1700

Most of what we now know as calculus—rules for differentiating and integrating elementary functions, solving simple differential equations, and expanding functions in power series—was known by the early eighteenth century and was included in the standard textbooks just mentioned. Nevertheless, there was much unfinished work. We list here a few of the open questions:

1. *Nonelementary Integrals.* Differentiation of elementary functions is an algorithmic procedure, and the derivative of any elementary function whatsoever, no matter how complicated, can be found if the investigator has sufficient patience. Such is not the

case for the inverse operation of integration. Many important elementary functions, such as $(\sin x)/x$ and e^{-x^2} , are not the derivatives of elementary functions. Since such integrals turned up in the analysis of some fairly simple motions, such as that of a pendulum, the problem of these integrals became pressing.

2. *Differential Equations.* Although integration had originally been associated with problems of area and volume, because of the importance of differential equations in mechanical problems the solution of differential equations soon became the major application of integration. The general procedure was to convert an equation to a form in which the derivatives could be eliminated by integrating both sides (reduction to quadratures). As these applications became more extensive, more and more cases began to arise in which the natural physical model led to equations that could not be reduced to quadratures. The subject of differential equations began to take on a life of its own, independent of the calculus.
3. *Foundational Difficulties.* The philosophical difficulties connected with the use of infinitesimal methods were paralleled by mathematical difficulties connected with the extension of the rules for operating with finite polynomials to infinite series. These difficulties were hidden for some time, and for a blissful century, mathematicians and physicists operated formally on power series as if they were finite polynomials. They did so even though it had been known since the time of Oresme that the partial sums of the harmonic series $1 + \frac{1}{2} + \frac{1}{3} + \dots$ grow arbitrarily large.

PROBLEMS AND QUESTIONS

Mathematical Problems

- 33.1.** The mathematical structures called *ordered fields*, have most of the properties of real numbers, in the sense that one can add, subtract, multiply, and divide them, as well as compare any two of them to determine which is the larger. One such field is formed by the real-valued *rational functions* of a real variable, that is, quotients of polynomials with real coefficients, an example of which is

$$r(x) = \frac{5x^2 - 7x + \sqrt{2}}{\pi x^3 - 46x^2 + 13}.$$

Two such functions $\frac{p(x)}{q(x)}$ and $\frac{P(x)}{Q(x)}$ are regarded as equal if $p(x)Q(x) = q(x)P(x)$ in the sense that the two sides are exactly the same polynomial, having exactly the same coefficients. Since polynomials have only a finite number of zeros, there is a largest zero for the numerator and denominator of a rational function $\frac{p(x)}{q(x)}$. For values of x larger than that largest zero, the fraction is of constant sign. We define $\frac{p(x)}{q(x)} > 0$ to mean that the values for all large x are positive, and $\frac{p(x)}{q(x)} > \frac{r(x)}{s(x)}$ to mean that $\frac{p(x)}{q(x)} - \frac{r(x)}{s(x)} > 0$.

Show that the rational function $f(x) = x$ is positive and larger than any constant function $g(x) = c$. Then show that no finite number of divisions of $f(x)$ by 2 will ever produce a function smaller than $g(x)$. Hence the rational functions are a *non-Archimedean ordered field*. (The standard real numbers are an Archimedean ordered field.)

- 33.2.** Show that the point at which the tangent to the curve $y = f(x)$ intersects the y axis is $(0, f(x) - xf'(x))$, and verify that the area under the curve $y = f(x) - xf'(x)$ from $x = 0$ to $x = a$ is twice the area between the curve $y = f(x)$ and the line $ay = f(a)x$. This result was used by Leibniz to illustrate the power of his infinitesimal methods.
- 33.3.** The principle that a curve is closely approximated by its tangent line at points near the point of tangency accounts for both the existence and the usefulness of the derivative. While the curve may have such a complicated equation that computations involving it are not feasible, the tangent line is computable, and computations on the tangent line involve only first-degree equations. That is the basis of Newton's method of approximating points where a function $f(x)$ is zero. You make a guess x_0 , compute the tangent line at the point $(x_0, f(x_0))$, which has equation $y - f(x_0) = f'(x_0)(x - x_0)$, and solve it for x when $y = 0$, getting a new guess x_1 , which one can hope is an improvement:

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

If x_1 still isn't good enough, repeat the process to get x_2 , and so on.

Use this technique to find $\sqrt{2}$. That is, find a zero of the function $f(x) = x^2 - 2$, for which $f'(x) = 2x$, starting with the guess $x_0 = 2$. What sequence of approximations do you obtain? How close is x_4 to $\sqrt{2}$?

Historical Questions

- 33.4.** When did Newton begin to create the calculus, and what problems did he solve with it?
- 33.5.** When did Leibniz begin to create the calculus, and what may have been his motive for doing so?
- 33.6.** What were the objections that philosophers raised against the techniques of calculus?

Questions for Reflection

- 33.7.** Just as Eudoxus solved the problem of incommensurables by making a definition of proportion to cover cases where no definition existed before, Newton's "theorem" asserting that quantities that approach each other monotonically and become arbitrarily close to each other in a finite time must become equal in an infinite time assumes that one has a definition of equality at infinity. Formulate such a definition.
- 33.8.** Draw a square and one of its diagonals. Then draw a very fine "staircase" by connecting short horizontal and vertical line segments in alternation, each segment crossing the diagonal. The total length of the horizontal segments is the same as the side of the square, and the same is true of the vertical segments, so that the total length of these segments is twice the length of a side. In an intuitive sense these segments do approximate the diagonal of the square, since they get closer and closer to it as the number of steps increases. This fact seems to imply that the diagonal of a square equals twice its side, which is absurd. Does this argument show that the method of indivisibles is wrong?

- 33.9.** In the passage quoted from the *Analyst*, Berkeley asserts that the experience of the senses provides the only foundation for our imagination. From that premise he concludes that we can have no understanding of infinitesimals. Analyze whether the premise is true, and if so, whether it implies the conclusion. Assuming that our thinking processes have been shaped by the evolution of the brain, for example, is it possible that some of our spatial and counting intuition is “hard-wired” and not the result of any previous sense impressions? The philosopher Immanuel Kant (1724–1804) thought so. Do we have the power to make correct judgments about spaces and times on scales that we have not experienced? What would Berkeley have said if he had heard Riemann’s argument that space may be finite, yet unbounded? If our intuition is “hard-wired,” does it follow that it is a perfectly accurate reflection of reality?

Consolidation of the Calculus

The calculus grew organically, sending forth branches while simultaneously putting down roots. The roots were the subject of philosophical speculation that eventually led to new mathematics as well, but the branches were natural outgrowths of pure mathematics that appeared very early in the history of the subject. In order to carry the story to a natural conclusion, we shall go beyond the time limits we have set for ourselves in this part and discuss results from the nineteenth century, but only in relation to calculus (analysis). The development of modern algebra, number theory, geometry, probability, and other subjects will be discussed in later chapters. In addition to the pioneers of calculus we have already discussed, we will be mentioning a number of outstanding eighteenth- and nineteenth-century mathematicians who made contributions to analysis, especially the following:

1. Leonhard Euler (1707–1783), a Swiss mathematician who became one of the early members of the Russian Academy of Sciences (1727–1741), then spent a quarter-century in Berlin (1741–1766) before returning to St. Petersburg when the Prussian Princess Catherine II (1762–1796) ruled there. He holds the record for having written the greatest volume of mathematical papers in all of history, amounting to more than 80 large volumes in the edition of his collected works. (A mathematician whose works fill 10 volumes is an extreme rarity.)
2. Jean le Rond d'Alembert (1717–1783), a French mathematician who made significant contributions to algebra, in which he attempted to prove that every polynomial with real coefficients can be written as a product of linear and quadratic factors with real coefficients. (If he had succeeded, he would as a by-product have proved the fundamental theorem of algebra.) He also contributed to partial differential equations (the vibrating string problem) and the foundations of mathematics. He was one of the authors of the great compendium of knowledge known as the *Encyclopédie*.
3. Joseph-Louis Lagrange (1736–1813), an Italian mathematician (Giuseppe-Luigi Lagrange), who spent most of his life in Berlin and Paris. He worked on many of the same problems in analysis as Euler. These two were remarkably prolific and between them advanced analysis, mechanics, and algebra immensely. Lagrange represented an algebraic point of view in analysis, generally eschewing appeals to geometry.
4. Adrien-Marie Legendre (1752–1833), a French mathematician who founded the theory of elliptic functions and made fundamental contributions to number

theory. He also was one of the earliest to recognize the importance of least-squares approximation.

5. Augustin-Louis Cauchy (1789–1856), the most prolific mathematician of the nineteenth century. He published constantly in the *Comptes rendus (Reports)* of the Paris Academy of Sciences. He raised the level of rigor in real analysis and was largely responsible for shaping one of three basic approaches to complex analysis. Although we shall be discussing some particular results of Cauchy in connection with the solution of algebraic and differential equations, his treatises on analysis are the contributions for which he is best remembered. He became a mathematician only after practicing as an engineer for several years.
6. Carl Gustav Jacob Jacobi (1804–1851), the first Jewish professor in Germany, who worked in many areas, including mechanics, elliptic and more general algebraic functions, differential equations, and number theory.
7. Karl Weierstrass (1815–1897), a professor at the University of Berlin from 1855 until his death. His insistence on clarity led him to reformulate much of analysis, algebra, and calculus of variations.
8. Bernhard Riemann (1826–1866), a brilliant geometer at the University of Göttingen. In frail health (he died young, of tuberculosis), he applied his wonderful intuition to invent a geometric style in complex analysis and algebra that complemented the analytic style of Weierstrass and the algebraic style of the Lagrangian tradition.

In our examination of the tree of calculus, we begin with the branches and will end with the roots.

34.1. ORDINARY DIFFERENTIAL EQUATIONS

Ordinary differential equations arose almost as soon as there was a language (differential calculus) in which they could be expressed. These equations were used to formulate problems from geometry and physics in the late seventeenth century, and the natural approach to solving them was to apply the integral calculus, that is, to reduce a given equation to quadratures. Leibniz, in particular, developed the technique now known as separation of variables as early as 1690 (Grosholz, 1987). In the simplest case, that of an ordinary differential equation of first order and first degree, one is seeking an equation $f(x, y) = c$, which may be interpreted as a conservation law if x and y are functions of time having physical significance. The conservation law is expressed as the differential equation

$$\frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy = 0.$$

The resulting equation is known as an exact differential equation, since the left-hand side is the exact differential of the function $f(x, y)$. To solve this equation, one has only to integrate the first differential with respect to x , adding an arbitrary function $g(y)$ to the solution, then differentiate with respect to y and compare the result with $\frac{\partial f}{\partial y}$ in order to get an equation for $g'(y)$, which can then be integrated.

If all equations were this simple, differential equations would be a very trivial subject. Unfortunately, it seems that nature tries to confuse us, multiplying these equations by arbitrary functions $\mu(x, y)$. That is, when an equation is written down as a particular case of a physical law, it often looks like

$$M(x, y) dx + N(x, y) dy = 0,$$

where $M(x, y) = \mu(x, y) \frac{\partial f}{\partial x}$ and $N(x, y) = \mu(x, y) \frac{\partial f}{\partial y}$, and no one can tell from looking at M just which factors in it constitute μ and which constitute $\frac{\partial f}{\partial x}$. To take the simplest possible example, the mass y of a radioactive substance that remains undecayed in a sample after time x satisfies the equation

$$dy - ky dx = 0,$$

where k is a constant. The mathematician's job is to get rid of $\mu(x, y)$ by looking for an "integrating factor" that will make the equation exact.¹ One integrating factor for this equation is $1/y$; another is e^{-kx} . (When the equation is solved, these are seen to be the same function.)

It appeared at a very early stage that finding an integrating factor is not in general possible, and both Newton and Leibniz were led to the use of infinite series with undetermined coefficients to solve such equations. Later, Maclaurin, was to warn against too hasty recourse to infinite series, saying that certain integrals could be better expressed geometrically as the arc lengths of various curves. But the idea of replacing a differential equation by a system of algebraic equations was very attractive. The earliest examples of series solutions were cited by Feigenbaum (1994). In his *Fluxions*, Newton considered the linear differential equation that we would now write as

$$\frac{dy}{dx} = 1 - 3x + x^2 + (1 + x)y.$$

Newton wrote it as $n/m = 1 - 3x + y + xx + xy$ and found that

$$y = x - x^2 + \frac{1}{3}x^3 - \frac{1}{6}x^4 + \frac{1}{30}x^5 - \frac{1}{45}x^6 - \dots$$

Similarly, in a paper published in the *Acta eruditorum* in 1693 (Gerhardt, 1971, Vol. 5, p. 287), Leibniz studied the differential equations for the logarithm and the arcsine in order to obtain what we now call the Maclaurin series of the logarithm, exponential, and sine functions. For example, he considered the equation $a^2 dy^2 = a^2 dx^2 + x^2 dy^2$ and assumed that $x = by + cy^3 + ey^5 + fy^7 + \dots$, thereby obtaining the series that represents the function $x = a \sin(y/a)$. Neither Newton nor Leibniz mentioned that the coefficients in these series were the derivatives of the functions represented by the series divided by the corresponding factorials. However, that realization came to John Bernoulli very soon after the publication

¹The equations presented in first courses on differential equations—those with variables separated, homogeneous equations, and linear equations—are precisely the equations for which an integrating factor is known.

of Leibniz' work. In a letter to Leibniz dated September 2, 1694 (Gerhardt, 1971, Vol. 3/1, p. 350), Bernoulli described essentially what we now call the Taylor series of a function. In the course of this description, he gave in passing what became a standard definition of a function, saying, "I take n to be a quantity formed in an arbitrary manner from variables and constants." Leibniz had used the word *function* as early as 1673, and in an article in the 1694 *Acta eruditorum* had defined a function to be "the portion of a line cut off by lines drawn using only a fixed point and a given point lying on a curved line." As Leibniz said, a given curve defines a number of functions: its abscissas, its ordinates, its subtangents, and so on. The problem that differential equations solve is to reconstruct the curve given the ratio between two of these functions.²

In classical terms, the solution of a differential equation is a function or family of functions. Given that fact, the ways in which a function can be presented become an important issue. With the modern definition of a function and the familiar notation, one might easily forget that in order to apply the theory of functions it is necessary to deal with particular functions, and these must be *presented* somehow. Bernoulli's description addresses that issue, although it leaves open the question of what methods of combining variables and constants are legal.

34.1.1. A Digression on Time

The Taylor series of a given function can be generated knowing the values of the function over any interval of the independent variable, no matter how short. Thus, a quantity represented by such a series is determined for all values of the independent variable when the values are given on any interval at all. Given that the independent variable is usually time, that property corresponds to physical determinacy: Knowing the full state of a physical quantity for some interval of time determines its values for all time. Lagrange, in particular, was a proponent of power series, for which he invented the term *analytic function*. However, as we now know, the natural domain of analytic function theory is the complex numbers. Now in mechanics the independent variable often represents time, and that fact raises an interesting question: Why should time be a complex variable? How do complex numbers turn out to be relevant to a problem where only real values of the variables have any physical meaning? To this question the eighteenth- and nineteenth-century mathematicians gave no answer. Indeed, it does not appear that they even asked the question very often. Extensive searches of the nineteenth-century literature by the present author have produced only the following comments on this interesting question, made by Weierstrass in 1885 (see his *Werke*, Bd. 3, S. 24):

It is very remarkable that in a problem of mathematical physics where one seeks an unknown function of two variables that, in terms of their physical meaning, can have only real values and is such that for a particular value of one of the variables the function must equal a prescribed function of the other, an expression often results that is an analytic function of the variable and hence also has a meaning for complex values of the latter.

²The mathematical meaning of the word *function* has always been somewhat at variance with its meaning in ordinary language. A person's function consists of the work the person does. Apparently, Leibniz pictured the curve as a means for producing these lines, which were therefore functions of the curve.

It is indeed very remarkable, but neither Weierstrass nor anyone since seems to have explained the mystery. Near the end of Weierstrass' life, Felix Klein (1897) remarked that if physical variables are regarded as complex, a rotating rigid body can be treated either as a motion in hyperbolic space or motion in Euclidean space accompanied by a strain. Perhaps, since they had seen that complex numbers were needed to produce the three real roots of a cubic equation, it may not have seemed strange to them that the complex-variable properties of solutions of differential equations are relevant in the study of problems generated by physical considerations involving only real variables. Time is sometimes represented as a two-dimensional quantity in connection with what are known as Gibbs random fields.

34.2. PARTIAL DIFFERENTIAL EQUATIONS

In the middle of the eighteenth century mathematical physicists began to consider problems involving more than one independent variable. The most famous of these is the vibrating string problem discussed by Euler, d'Alembert, and Daniel Bernoulli (1700–1782, son of John Bernoulli) during the 1740s and 1750s.³ This problem led to the one-dimensional wave equation

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2},$$

with the initial conditions $u(x, 0) = f(x)$, $\frac{\partial u}{\partial t}(x, 0) = 0$. Here $u(x, t)$ is the height of the point of the string above x at time t . Daniel Bernoulli solved this equation in the form of an infinite double trigonometric series

$$u(x, t) = \sum_{n=1}^{\infty} a_n \sin nx \cos nct,$$

claiming that the a_n could be chosen so that $\sum_{n=1}^{\infty} a_n \sin nx = f(x)$. This solution was criticized by Euler, leading to a debate over the allowable methods of defining functions and the proper definition of a function.

The developments that grew out of trigonometric-series techniques like this one by Daniel Bernoulli will be discussed in Chapter 42, along with the development of real analysis in general. For the rest of the present section, we confine our discussion to power-series techniques of solving partial differential equations.

In the nineteenth century, Newton's power-series method was applied to the heat equation

$$\frac{\partial u}{\partial t} = a \frac{\partial^2 u}{\partial x^2}$$

³The problem had been considered a generation earlier by Brook Taylor, who made the assumption that the restoring force on the string at any point and any time was proportional to the curvature of its shape at that point and time. Since the curvature is essentially the second derivative with respect to arc length, this condition, when linearized, amounts to the partial differential equation used by d'Alembert.

by Joseph Fourier, who is actually better known for applying trigonometric series and integrals in such cases. (In fact, they are called Fourier series and integrals in his honor.) In this equation, $u(x, t)$ represents the temperature at time t at point x in a long thin wire. Assuming that the temperature at x at time $t = 0$ is $\varphi(x)$ and $a = 1$, Fourier obtained the solution

$$u(x, t) = \sum_{r=0}^{\infty} \frac{\varphi^{(2r)}(x)}{r!} t^r.$$

As it turns out, this series often diverges for all nonzero values of t .

It was not until the nineteenth century that mathematicians began to worry about the convergence of series solutions. First Cauchy, and then Weierstrass produced proofs that the series do converge for *ordinary* differential equations, provided that the coefficients have convergent series representations. For *partial* differential equations, between 1841 and 1876, Cauchy, Jacobi, Weierstrass, Weierstrass' student Sof'ya Kovalevskaya (1850–1891), and Gaston Darboux (1842–1917), produced theorems that guaranteed convergence of the formally generated power series. In general, however, it turned out that the series formally satisfying the equation could actually diverge, and that the algebraic form of the equation controlled whether it did or not. Kovalevskaya showed that in general the power series solution for the heat equation diverges if the initial temperature distribution is prescribed, even when that temperature is an analytic function of position. (This is the case considered by Fourier.) She showed, however, that the series converges if the temperature and temperature gradient at one point are prescribed as analytic functions of time. More generally, she showed that the power-series solution of any initial-value problem in “normal form” would converge. Normal form is relative to a particular variable that occurs in the equation. It means that the initial conditions are imposed on a variable whose highest-order pure derivative in the equation equals the order of the equation. The heat equation is in normal form relative to the spatial variable, but not relative to the time variable.

34.3. CALCULUS OF VARIATIONS

The notion of function lies at the heart of calculus. The usual picture of a function is of one *point* being mapped to another *point*. However, the independent variable in a function can be a curve or surface as well as a point. For example, given a curve γ that is the graph of a function $y = f(x)$ between $x = a$ and $x = b$, we can define its length as

$$\Lambda(\gamma) = \int_a^b \sqrt{1 + (f'(x))^2} dx.$$

One of the important problems in the history of geometry has been to pick out the curve γ that minimizes $\Lambda(\gamma)$ and satisfies certain extra conditions, such as joining two fixed points P and Q on a surface or enclosing a fixed area A . The calculus technique of “setting the derivative equal to zero” needs to be generalized for such problems, and the techniques for doing so constitute the calculus of variations. The history of this outgrowth of the calculus

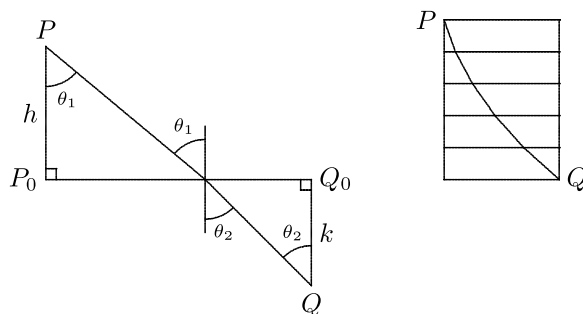


Figure 34.1. Left: Fermat's principle. The time of travel from P to Q is a minimum if the ray crosses the interface at the point where $\sin(\theta_1)/v_1 = \sin(\theta_2)/v_2$. Right: Application of this principle to the brachistochrone, assuming the speed varies continuously in proportion to the square root of the distance of descent.

has been studied in many classic works, such as those by Woodhouse (1810),⁴ Todhunter (1861), and Goldstine (1980), and in articles like the one by Kreyszig (1993).

As with the ordinary calculus, the development of calculus of variations proceeded from particular problems solved by special devices to general techniques and algorithms based on theoretical analysis and rigorous proof. In the seventeenth century there were three such special problems that had important consequences. The first was the brachistochrone (shortest-time) problem for an object crossing an interface between two media while moving from one point to another. In the simplest case (Fig. 34.1), the interface is a straight line, and the time required to travel from P to Q at speed v_1 above the line P_0Q_0 and speed v_2 below it is to be minimized. If the two speeds are not the same, it is clear that the path of minimum time will not be a straight line, since time can be saved by traveling a slightly longer distance in the medium in which the speed is greater. The path of minimum time turns out to be the one in which the sines of the angle of incidence and refraction have a fixed ratio, namely the ratio of the speeds in the two media. (Compare this result with the shortest reflected path in a single medium, discussed in Problem 15.1 of Chapter 15, which is also a path of minimum time.)

Fermat's principle, which asserts that the path of a light ray is the one that requires least time, found application in the second problem, stated as a challenge by John Bernoulli in 1696: *Find the curve down which a frictionless particle will slide from point P to point Q under the influence of gravity in minimal time.* Since the speed of a falling body is proportional to the square root of the distance fallen, Bernoulli reasoned that the sine of the angle between the tangent and the vertical would be proportional to the square root of the

⁴The treatise of Woodhouse is a textbook as much as a history, and its last chapter is a set of 29 examples posed as exercises for the reader with solutions provided. The book also marks an important transition in British mathematics. Woodhouse says in the preface that, "In a former Work, I adopted the foreign notation. . .". The foreign notation was the Leibniz notation for differentials, in preference to the dot above the letter that Newton used to denote his fluxions. He says that he found this notation even more necessary in calculus of variations, since he would otherwise have had to adopt some new symbol for Lagrange's variation. But he then goes on to marvel that Lagrange had taken the reverse step of introducing Newton's fluxion notation into the calculus of variations.

vertical coordinate, assuming the vertical axis directed downward.⁵ In that way, Bernoulli arrived at a differential equation for the curve:

$$\frac{dy}{dx} = \sqrt{\frac{y}{a-y}}.$$

Here we have taken y as the vertical coordinate, directed downward. He recognized this equation as the differential equation of a cycloid and thus concluded that this curve, which Christiaan Huygens (1629–1695) had studied because it enabled a clock to keep theoretically perfect time (the tautochrone property, discussed in Chapter 39), also had the brachistochrone property. The challenge problem was solved by Bernoulli himself, by his brother James, and by both Newton and Leibniz.⁶ According to Woodhouse (1810, p. 150), Newton's anonymously submitted solution was so concise and elegant that John Bernoulli knew immediately who it must be from. He wrote, "Even though the author, from excessive modesty, does not give his name, we can nevertheless tell certainly by a number of signs that it is the famous Newton; and even if these signs were not present, seeing a small sample would suffice to recognize him, as *ex ungue Leonem*."⁷

The third problem, that of finding the cross-sectional shape of the optimally streamlined body moving through a resisting medium, is discussed in the scholium to Proposition 34 (Theorem 28) of Book 2 of Newton's *Principia*.

34.3.1. Euler

Variational problems were categorized and systematized by Euler in a large treatise in 1744 named *Methodus inveniendi lineas curvas (A Method of Finding Curves)*. In this treatise Euler set forth a series of problems of increasing complexity, each involving the finding of a curve having certain extremal properties, such as minimal length among all curves joining two points on a given surface.⁸ Proposition 3 in Chapter 2, for example, asks for the minimum value of an integral $\int Z dx$, where Z is a function of variables, x , y , and $p = y' = \frac{dy}{dx}$. Based on his previous examples, Euler derived the differential equation

$$0 = N dx - dP,$$

where $dZ = M dx + N dy + P dp$ is the differential of the integrand Z . Since $N = \frac{\partial Z}{\partial y}$ and $P = \frac{\partial Z}{\partial p}$, this equation could be written in the form that is now the basic equation of the

⁵As discussed in Chapter 27, the Muslim scholars ibn Sahl and al-Haytham knew that the ratio of the sines of the angles of incidence and refraction was constant at a point where two media meet. The Europeans Thomas Harriot, Willebrod Snell, and René Descartes derived the law of refraction from theoretical principles and deduced that the ratio of these sines is the ratio of the speeds of propagation in the two media. Fermat's principle, which was stated in a letter written in 1662, uses this law to show that the time of travel from a point in one medium to a point in the other is minimal.

⁶Newton apparently recognized structural similarities between this problem and his own optimal-streamlining problem (see Goldstine, 1980, pp. 7–35).

⁷A Latin proverb much in vogue at the time. It means literally "from [just] the claw [one can recognize] the Lion."

⁸This problem was Example 4 in Chapter 4 of the treatise.

calculus of variations, and is known as Euler's equation:

$$\frac{\partial Z}{\partial y} = \frac{d}{dx} \left(\frac{\partial Z}{\partial y'} \right).$$

In Chapter 3, Euler generalized this result by allowing Z to depend on additional parameters and applied his result to find minimal surfaces. In an appendix he studied elastic curves and surfaces, including the problem of the vibrating membrane. This work was being done at the very time when Euler's former colleague Daniel Bernoulli was studying the simpler problem of the vibrating string. In a second appendix, Euler showed how to derive the equations of mechanics from variational principles, thus providing a unifying mathematical principle that applied to both optics (Fermat's principle) and mechanics.⁹

34.3.2. Lagrange

The calculus of variations acquired "variations" and its name as the result of a letter written by Lagrange to Euler in 1755. In that letter, Lagrange generalized Leibniz' differentials from points to curves, using the Greek δ instead of the Latin d to denote them. Thus, if $y = f(x)$ was a curve, its *variation* δy was a small perturbation of it. Just as dy was a small change in the value of y at a point, δy was a small change in all the values of y at all points. The variation operator δ can be manipulated quite easily, since it commutes with differentiation and integration: $\delta y' = (\delta y)'$ and $\delta \int Z dx = \int \delta Z dx$. With this operator, Euler's equation and its many applications were easy to derive. Euler recognized the usefulness of what Lagrange had done and gave the new theory the name it has borne ever since: calculus of variations.

Lagrange also considered extremal problems with constraint and introduced the famous Lagrange multipliers as a way of turning these relative (constrained) extrema into absolute (unconstrained) extrema. Euler had given an explanation of this process earlier. Woodhouse (1810, p. 79) thought that Lagrange's systematization actually deprived Euler's ideas of their simplicity.

34.3.3. Second-Variation Tests for Maxima and Minima

Like the equation $f'(x) = 0$ in calculus, the Euler equation is only a necessary condition for an extremal, not sufficient, and it does not distinguish between maximum, minimum, and neither. In general, however, if Euler's equation has only one solution, and there is good reason to believe that a maximum or minimum exists, the solution of the Euler equation provides a basis to proceed in practice. Still, mathematicians were bound to explore the question of distinguishing maxima from minima. Such investigations were undertaken by Lagrange and Legendre in the late eighteenth century.

In 1786 Legendre was able to show that a sufficient condition for a minimum of the integral

$$I(y) = \int_a^b f(x, y, y') dx,$$

⁹ One of his results is that a particle moving over a surface and free of any forces tangential to the surface will move along a geodesic of that surface. One cannot help seeing in this result an anticipation of the basic principle of general relativity (see Chapter 39 below).

at a function satisfying Euler's necessary condition, was $\frac{\partial^2 f}{\partial y'^2} > 0$ for all x and that a sufficient condition for a maximum was $\frac{\partial^2 f}{\partial y'^2} < 0$.

In 1797 Lagrange published a comprehensive treatise on the calculus, in which he objected to some of Legendre's reasoning, noting that it assumed that certain functions remained finite on the interval of integration (Dorofeeva, 1998, p. 209).

34.3.4. Jacobi: Sufficiency Criteria

The second-variation test is strong enough to show that a solution of the Euler equation really is an extremal among the smooth functions that are "nearby" in the sense that their values are close to those of the solution and their derivatives also take values close to those of the derivative of the solution. Such an extremal was called a *weak extremal* by Adolf Kneser (1862–1930). Jacobi had the idea of replacing the curve $y(x)$ that satisfied Euler's equation with a family of such curves depending on parameters (two in the case we have been considering) $y(x, \alpha_1, \alpha_2)$ and replacing the nearby curves $y + \delta y$ and $y' + \delta y'$ with values corresponding to different parameters. In 1837—see Dorofeeva (1998) or Fraser (1993)—he finally solved the problem of finding sufficient conditions for an extremal. He included his solution in the lectures on dynamics that he gave in 1842, which were published in 1866, after his death. The complication that had held up Jacobi and others was the fact that sometimes the extremals with given endpoints are not unique. The most obvious example is the case of great circles on the sphere, which satisfy the Euler equations for the integral that gives arc length subject to fixed endpoints. If the endpoints happen to be antipodal points, all great circles passing through the two points have the same length. Weierstrass was later to call such pairs of points *conjugate points*. Jacobi gave a differential equation whose solutions had zeros at these points and showed that Legendre's criterion was correct, provided that the interval $(a, b]$ contained no points conjugate to a .

34.3.5. Weierstrass and his School

A number of important advances in the calculus of variations were due to Weierstrass, such as the elimination of some of the more restrictive assumptions about differentiability and taking account of the distinction between a lower bound and a minimum.¹⁰

An important example in this connection was Riemann's use of *Dirichlet's principle* to prove the Riemann mapping theorem, which asserts that any simply connected region in the plane except the plane itself can be mapped conformally onto the unit disk $\Delta = \{(x, y) : x^2 + y^2 < 1\}$. That principle required the existence of a real-valued function $u(x, y)$ that minimizes the integral

$$\iint_{\Delta} \left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial y} \right)^2 dx dy$$

among all functions $u(x, y)$ taking prescribed values on the boundary of the disk. That function is the unique harmonic function¹¹ in Δ with the given boundary values. In 1870,

¹⁰This distinction was pointed out by Gauss as early as 1799, in his criticism of d'Alembert's 1746 proof of the fundamental theorem of algebra.

¹¹A brief definition of a harmonic function is that its graph is the surface of a nonvibrating flexible membrane.

Weierstrass called attention to the integral

$$\Phi(\varphi) = \int_{-1}^{+1} x^2 (\varphi'(x))^2 dx,$$

which when combined with the boundary condition $\varphi(-1) = a$, $\varphi(+1) = b$, can be made arbitrarily small by taking k sufficiently large in the formula

$$\varphi(x) = \frac{a+b}{2} + \frac{b-a}{2} \frac{\arctan(kx)}{\arctan(k)},$$

yet (if $a \neq b$) cannot be zero for any function φ satisfying the boundary conditions and such that φ' exists at every point.

Weierstrass' example was a case where it was necessary to look outside the class of smooth functions for a minimum of the functional. The limiting position of the graphs of the functions for which the integral approximates its minimum value consists of the two horizontal lines from $(-1, a)$ to $(0, a)$, from $(0, b)$ to $(+1, b)$, and the section of the y -axis joining them (see Fig. 34.2).

Weierstrass thought of the smoothness assumptions as necessary evils. He recognized that they limited the generality of the results, yet he saw that without them no application of the calculus was possible. The result is a certain vagueness about the formulation of minimal principles in physics. A certain functional must be a minimum *assuming* that all the relevant quantities are differentiable a sufficient number of times. Obviously, if a functional can be extended to a wider class of functions in a natural way, the minimum reached may be smaller, or the maximum larger. To make the restrictions as weak as possible, Weierstrass imposed the condition that the partial derivatives of the integrand should be continuous at corners. An extremal among all functions satisfying these less restrictive hypotheses was

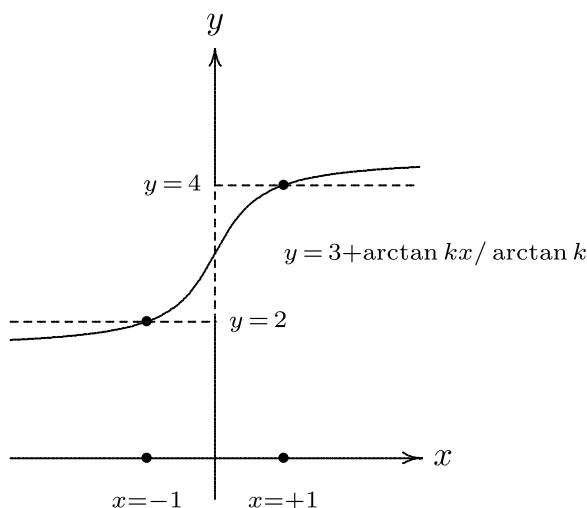


Figure 34.2. The functional $\Phi(y) = \int_{-1}^{+1} (xy'(x))^2 dx$ does not assume its minimum value for continuously differentiable functions $y(x)$ satisfying $y(-1) = 2$, $y(+1) = 4$. The limiting position of a minimizing sequence is the dashed line.

called a *strong* extremal. The corner condition was also found in 1877 by G. Erdmann (dates unknown), a teacher at the Gymnasium in Königsberg, who proved in 1878 that Jacobi's sufficient condition for a weak extremal was also necessary.

34.4. FOUNDATIONS OF THE CALCULUS

The British and Continental mathematicians both found the power of the calculus so attractive that they applied and developed it (sending forth new branches), all the while struggling to be clear about the principles they were using (extending its roots). The branches grew more or less continuously from the beginning. The development of the roots was slower and more sporadic. A satisfactory consensus was achieved only late in the nineteenth century, with the full development of real analysis.

The source of the difficulty was the introduction of the infinite into analysis in the form of infinitesimal reasoning. As mentioned in the previous chapter, Leibniz believed in actual infinitesimals, levels of magnitude that were real, not zero, but so small that no accumulation of them could ever exceed any finite quantity. His dx was such an infinitesimal, and a product of two, such as $dx dy$ or dx^2 , was a higher-order infinitesimal, so small that no accumulation of such could ever exceed any infinitesimal of the first order. On this view, even though theorems established using calculus were not absolutely accurate, the errors were below the threshold of human perception and therefore could not matter in practice. Newton was probably alluding to this belief of Leibniz when, in his discussion of the quadrature of curves (1704), he wrote, "In rebus mathematicis errores quam minimi non sunt contemnendi" ("Errors, no matter how small, are not to be allowed in mathematics").¹²

Newton knew that his arguments could have been phrased using the Eudoxan method of exhaustion. In his *Principia* he wrote that he used his method of first and last ratios "to avoid the tediousness of deducing involved demonstrations *ad absurdum*, according to the method of the ancient geometers." That is to say, to avoid the trichotomy arguments used by Archimedes.

There seemed to be three approaches that would allow the operation that we now know as integration to be performed by antidifferentiation of tangents. One is the infinitesimal approach of Leibniz, characterized by Mancosu (1989) as "static." That is, a tangent is a state or position of a line, namely that of passing through two infinitely near points. The second is Newton's "dynamic" approach, in which a fluxion is the velocity of a moving object. The third is the ancient method of exhaustion. In principle, a reduction of calculus to the Eudoxan theory of proportion is possible. Psychologically, it would involve not only a great deal of tedium, as Newton noted, but also a great deal of confusion. If mathematicians had been shackled by the requirements of this kind of rigor, the amount of geometry and analysis created would have been much smaller than it was.

In the eighteenth century, however, better expositions of the calculus were produced by d'Alembert and others. In his article on the differential for the famous *Encyclopédie*, d'Alembert wrote that $0/0$ could be equal to anything, and that the derivative $\frac{dy}{dx}$ was not actually 0 divided by 0, but the limit of finite quotients as numerator and denominator tended to zero. (This was essentially what Newton had said in his *Principia*.)

¹²As we saw in the last chapter, Berkeley flung these very words back at Newton.

34.4.1. Lagrange's Algebraic Analysis

The attempt to be clear about infinitesimals or to banish them entirely took many forms during the eighteenth and nineteenth centuries. One of them (see Fraser, 1987) was Lagrange's exposition of analytic functions. Lagrange understood the term *function* to mean a formula composed of symbols representing variables and arithmetic operations. He argued that "in general" (with certain obvious exceptions) every function $f(x)$ could be expanded as a power series, based on Taylor's theorem, for which he provided his own form of the remainder term. He claimed that the hypothetical expansion

$$\sqrt{x+h} = \sqrt{x} + ph + qh^2 + \cdots + h^{m/n}$$

could not occur, since the left-hand side has only two values, while the right-hand side has n values.¹³ In this way, he ruled out fractional exponents. Negative exponents were ruled out by the mere fact that the function was defined at $h = 0$. The determinacy property of analytic functions was used implicitly by Lagrange when he assumed that any zero of a function must have finite order, as we would say (Fraser, 1987, p. 42).

The advantage of confining attention to functions defined by power series is that the derivative and integral of such a function have a perfectly definite meaning. Lagrange advocated it on the grounds that it showed the qualitative difference between the functions dx and x .

34.4.2. Cauchy's Calculus

The modern presentation of calculus owes a great deal to the textbooks of Cauchy, written for his lectures at the Ecole Polytechnique during the 1820s. Cauchy recognized that calculus could not get by without something equivalent to infinitesimals. He defined a function $f(x)$ to be continuous if the absolute value of the difference $f(x + \alpha) - f(x)$ "decreases without limit along with that of α ." He continues:

In other words, *the function $f(x)$ remains continuous with respect to x in a given interval, if an infinitesimal increase in the variable within this interval always produces an infinitesimal increase in the function itself.*

Cauchy did not discuss the question whether only one single point x is being considered or the increase is being thought of as occurring at all points simultaneously. It turns out that the *size* of the infinitesimal change in $f(x)$ corresponding to a given change in x may vary from one point to another and from one function to another. Stronger assumptions, invoking the concepts of uniform continuity and equicontinuity are needed to guarantee results such as Cauchy stated here. In particular, he uniform convergence and continuity but did not say so. Cauchy defined a limit in terms of the "successive values attributed to a variable," approaching a fixed value and ultimately differing from it by an arbitrarily small amount. This definition can be regarded as an informal version of what we now state precisely with deltas and epsilons; and Cauchy is generally regarded, along with Weierstrass, as one of the

¹³This kind of reasoning was used by Abel in the nineteenth century to prove that there is no finite algebraic algorithm for solving the general equation of degree 5.

people who finally made the foundations of calculus secure. Yet Cauchy's language clearly presumes that infinitesimals are real. As Laugwitz (1987, p. 272) says:

All attempts to understand Cauchy from a 'rigorous' theory of real numbers and functions including uniformity concepts have failed... One advantage of modern theories like the Non-standard Analysis of Robinson... [which includes infinitesimals] is that they provide consistent reconstructions of Cauchy's concepts and results in a language which sounds very much like Cauchy's.

The secure foundation of modern analysis owes much to Cauchy's treatises. As Grabiner (1981) said, he applied ancient Greek rigor and modern algebraic techniques to derive results from analysis.

PROBLEMS AND QUESTIONS

Mathematical Problems

- 34.1.** Consider the one-dimensional heat equation, according to which the temperature u at point x along a line (say a wire) at time t satisfies

$$\frac{\partial u}{\partial t} = k \frac{\partial^2 u}{\partial x^2},$$

where k is a constant of proportionality. Assume the units of time and distance are chosen so that $k = 1$. If the initial temperature distribution is given by the so-called *witch of Agnesi*¹⁴ $u(x, 0) = (1 + x^2)^{-1}$ (so that the temperature has some resemblance to a bell-shaped curve), assume that

$$u(x, t) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} c_{mn} x^m t^n.$$

Use the fact that

$$u(x, 0) = 1 - x^2 + x^4 - x^6 + \dots$$

for all small x to conclude that

$$c_{m0} = \begin{cases} 0, & \text{if } m \text{ is odd,} \\ (-1)^p, & \text{if } m = 2p. \end{cases}$$

¹⁴In her calculus textbook, Maria Gaetana Agnesi called this curve *la versiera*, meaning *twisted*. It was incorrectly translated into English, apparently because of the resemblance of this word to *l'avversiera*, meaning *wife of the Devil*.

Then differentiate formally, and show that the assumed series for $u(x, t)$ must be

$$u(x, t) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} (-1)^{m+n} \frac{(2m+2n)!}{(2m)!n!} x^{2m} t^n.$$

Show that this series diverges for all nonzero values of t when $x = 0$.

- 34.2.** There are yet more subtleties in the notion of continuity than even Cauchy realized. In one of his works, he had stated the theorem that the sum of a series of continuous functions is continuous. Abel, who admired Cauchy's mathematics (while regarding Cauchy himself as rather crazy), diplomatically pointed out that "this theorem appears to admit some exceptions." In fact,

$$\sum_{n=1}^{\infty} \frac{1}{n} \sin nx = \begin{cases} +\frac{\pi-x}{2} & \text{if } 0 < x < \pi, \\ 0 & \text{if } x = k\pi, \quad k = 0, \pm 1, \pm 2, \dots \\ -\frac{\pi-x}{2} & \text{if } -\pi < x < 0. \end{cases}$$

Since Cauchy had argued that an infinitesimal change in x will produce an infinitesimal change in each term $\frac{\sin nx}{n}$, why does an infinitesimal increase in x starting at $x = 0$ not produce an infinitesimal change in the sum of this series?

- 34.3.** Fill in the details of Weierstrass' example of a functional that does not assume its minimum value subject to certain endpoint conditions. In Fig. 34.2, the function $y_k = 3 + \arctan(kx)/\arctan(k)$ satisfies the endpoint conditions that $y(-1) = 2$ and $y(+1) = 4$. Using partial fractions to do the integration, you can show that

$$\int_{-1}^{+1} (xy'_k(x))^2 dx = \left(\frac{1}{k} - \frac{1}{(1+k^2)\arctan(k)} \right),$$

which obviously tends to zero as $k \rightarrow \infty$. For the functional actually to be zero, however, $y'(x)$ would have to be identically zero except at $x = 0$, and so $y(x)$ would have to be 2 for $x < 0$ and 4 for $x > 0$.

Historical Questions

- 34.4.** How does the calculus of variations differ from ordinary calculus?
- 34.5.** What new methodological questions arose in the course of solving the problem of the vibrating string?
- 34.6.** What solutions did nineteenth-century analysts like Cauchy and Weierstrass find to the philosophical difficulties connected with infinitesimals?

Questions for Reflection

- 34.7.** Is it possible to make calculus "finitistic," so that each step in its development refers only to a finite number of concrete things? Or is the infinite inherent in the subject? In particular, does Lagrange's approach, developing functions as power series

and defining the derivative as the coefficient of the first-degree term, satisfy such a requirement and eliminate the need for infinitesimals?

- 34.8.** What sense can you make out of time as a complex variable? If it has no meaning at all, why did Weierstrass and his students think it important to use complex variables in solving differential equations?
- 34.9.** What differences are there between an algebraic equation and a differential equation? What does the term *solution* mean for each of them?

SPECIAL TOPICS

The ordering of material that we have used up to now—first, by cultures and, within each culture, roughly chronological—becomes useless after the beginning of the eighteenth century. From that point on, there is essentially only one mathematical culture, a world-wide one, with a broad consensus as to methods, although some specialties are more concentrated in one geographical area than another. As for chronology, so much mathematics has been produced every year, and mathematics has been advancing along so many broad fronts, that a chapter devoted to a single decade in the eighteenth century or a single year in the twentieth would be prodigiously long. As the time period grew shorter and the chapters grew longer, all perspective would be lost. For that reason, this final part of the history, except for Chapter 35, which discusses women mathematicians in the late nineteenth and early twentieth centuries, consists of chapters, each of which is devoted to the development of a single subject area.

In an effort to convey as much mathematics as possible in this book, we have slighted some other questions of sociological and political interest, such as the increasing democratization of mathematics that accompanied the increase in prosperity after the industrial revolution, its opening up to people from working-class backgrounds. Especially important in that democratization was the gradual involvement of women in the mathematical world. We shall devote the first chapter in this final part to that subject. For lack of space, we are forced to omit other interesting subjects, such as the influence of the Nazi and Communist regimes on mathematics in Germany and the Soviet Union and the impact of the Cold War on mathematical research in the United States.

We ended our narrative of the development of different mathematical subjects at different points. We left the story of both algebra and geometry at the point they had reached around the beginning of the seventeenth century, and we left the story of calculus and its outgrowths in the nineteenth century. Certain prominent parts of mathematics, such as probability, mathematical logic, set theory, and modern number theory have hardly been mentioned at all. While the enormous literature generated by these subjects in the modern era makes the task of summarizing them nearly impossible, we can at least make a grand sweep of each of them to provide some measure of completeness to our coverage of the world of mathematics. These last eleven chapters will fill in some of these gaps. These chapters, much more than those that have preceded, are written in the style that we called *heritage* in Chapter 1. That is, they aim to show how certain familiar features of modern mathematics arose rather than to describe objectively what mathematical life was like in the past.

Contents of Part VII

1. Chapter 35 (Women Mathematicians), as mentioned above, discusses women mathematicians in the late nineteenth and early twentieth centuries.
2. Chapter 36 (Probability) traces the history of probability from the Renaissance to the nineteenth century.
3. Chapter 37 (Algebra from 1600 to 1850) discusses the development of algebra up to the mid-nineteenth century.
4. Chapters 38–40 (Projective and Algebraic Geometry and Topology, Differential Geometry, Non-Euclidean Geometry) describe, as their titles indicate, the development of projective and algebraic geometry, differential geometry, and non-Euclidean geometry, respectively, to the end of the nineteenth century.
5. Chapter 41 (Complex Analysis) is devoted to complex analysis.
6. Chapters 42 and 43 (Real Numbers, Series, and Integrals; Foundations of Real Analysis) describe the parts of real analysis as branches and roots of calculus, pursuing the analogy introduced in Chapter 34.
7. Chapter 44 (Set Theory) discusses the origin and development of set theory from the 1880s through the early twentieth century.
8. Chapter 45 (Logic) discusses mathematical logic and the philosophy of mathematics from the mid-nineteenth century through the mid-twentieth century.

Women Mathematicians

The history of women's participation in mathematical research has become an area of considerable interest over the past few decades. The movement of women into mathematics blossomed enormously during the late twentieth century, the result of a long and arduous struggle by brave and determined pioneers. Unfortunately for those who write the history of mathematics, almost all of this movement occurred after the time period one can reasonably cover in a single semester. In the preceding chapters, only three women—Hypatia, Maria Gaetana Agnesi, and Sof'ya Kovalevskaya—are prominent enough to merit mention. Hypatia was primarily a philosopher, and the details of her mathematical activity are not known. What Agnesi and Kovalevskaya did is well understood and appreciated. However, they enter the picture, as we have seen, near the end of the time period we are covering. To make up in some degree for these omissions, we discuss here three of the women who, in the years between 1850 and 1935, made a mark on the mathematical world in their time, overcoming prejudice and personal hardship in many cases in order to do so.

Women first began to break into the intellectual world of modern Europe in the eighteenth century, mingling with the educated society of their communities, but not allowed to attend the meetings of scientific societies. The struggle for a woman's right to be a scientist or mathematician was very much an obstacle course, similar to running the high hurdles. The first hurdle was to get the family to support a scientific education. That hurdle alone caused many to drop out at the very beginning, leaving only a few lucky or very determined women to go on to the second hurdle, gaining access to higher education. The second hurdle began to be crossed in the late nineteenth century. On the continent, a few women were admitted to university lectures without being matriculated, as exceptional cases. These cases established a precedent, and the exceptions eventually became regularized. In Britain, the University of London began admitting women in the 1870s, and in the United States there were women's colleges for undergraduate education. The opening of Bryn Mawr College in 1885 with a program of graduate studies in mathematics was an important milestone in this progress. Once a woman had gotten past the second hurdle, the third and highest of all had to be faced: getting hired and accepted by the mathematical community, and finding time to do mathematics in addition to the heavy familial responsibilities laid on women by society. The three pioneers we are about to discuss had to improvise their solutions to this problem. The fundamental societal changes needed to provide women with

the same assured, routine access that men enjoyed when pursuing such a career required many decades to be recognized and partially implemented.

35.1. SOF'YA KOVALEVSKAYA

Most of the early women mathematicians came from a leisured class of people with independent incomes. Only such people can afford both to defy convention and to spend most of their time pursuing what interests them. However, merely having an independent income was not in itself sufficient to draw a young woman into a scientific career. In most cases, some contact with intellectual circles was present as well. Hypatia was the daughter of a distinguished scholar, and Maria Gaetana Agnesi's father encouraged her by hiring tutors to instruct her in classical languages. In the case of Sof'ya Kovalevskaya, the urge to study mathematics and science fused with her participation in the radical political and social movements of her time, which looked to science as the engine of material progress and aimed to establish a society in accordance with the ideals of democracy and socialism.

She was born Sof'ya Vasil'evna Kryukovskaya in Moscow, where her father was an officer in the army, on January 15, 1850 (January 3 on the Julian calendar in effect in the Russia of her day). As a child she looked with admiration on her older sister Anna (1843–1887) and followed Anna's lead into radical political and social activism. According to her Polish tutor, she showed talent for mathematics when still in her early teens. She also showed great sympathy for the cause of Polish independence during the rebellion of 1863, which was crushed by the Tsar's troops. When she was 15, one of her neighbors, a physicist, was impressed upon discovering that she had invented the rudiments of trigonometry all by herself in order to read a book on optics; he urged her father to allow her to study more science. She was allowed to study up through the beginnings of calculus with a private tutor in St. Petersburg, but matriculation at a Russian university did not appear to be an option. Thinking that Western Europe was more enlightened in this regard, many young Russian women used a variety of methods to travel abroad. Some were able to persuade their parents to let them go. Others had to adopt more radical means, either running away or arranging a marriage of convenience, in Sof'ya's case to a young radical publisher named Vladimir Onufrevich Kovalevskii (1842–1883). They were married in 1869 and soon after left for Vienna and Heidelberg, where Kovalevskaya studied science and mathematics for a year without being allowed to enroll in the university, before moving on to Berlin with recommendations from her Heidelberg professors to meet the man who was to have the dominant influence on her professional life, Weierstrass. At Berlin also, the university would not accept her as a regular student, but Weierstrass agreed to tutor her privately.

Although the next four years were extremely stressful for a number of personal reasons, her regular meetings with Weierstrass brought her knowledge of mathematical analysis up to the level of the very best students in the world (those attending Weierstrass' lectures). By 1874, Weierstrass thought she had done more than enough work for a degree and proposed three of her papers as dissertations. Since Berlin would not award the degree, he wrote to the University of Göttingen and requested that the degree be granted *in absentia*. It was, and one of the three papers became a classic work in differential equations, published the following year in the most distinguished German journal, the *Journal für die reine und angewandte Mathematik*.

The next eight years may well be described as Kovalevskaya's wandering in the intellectual wilderness. She and Vladimir, who had obtained a doctorate in geology from the

University of Jena, returned to Russia; but neither found an academic position commensurate with their talents. They began to invest in real estate, in the hope of gaining the independent wealth they would need to pursue their scientific interests. During this time, Kovalevskaya gave birth to a daughter, Sof'ya Vladimirovna Kovalevskaya (1878–1951). Soon afterward, their investments failed, and they were forced to declare bankruptcy. Vladimir's life began to unravel at this point; and Kovalevskaya, knowing that she would have to depend on herself, reopened her mathematical contacts and began to attend mathematical meetings. Recognizing the gap in her résumé since her dissertation, she asked Weierstrass for a problem to work on in order to reestablish her credentials. While she was in Paris in the spring of 1883, Vladimir (back in Russia) committed suicide, leading Sof'ya to an intense depression that nearly resulted in her own death. When she recovered, she resumed work on the problem that Weierstrass had given her. Meanwhile, Weierstrass and his student Gösta Mittag-Leffler (1846–1927) collaborated to find her a teaching position at the newly founded institution in Stockholm.¹ At first she was *Privatdozent*, meaning that she was paid a certain amount for each student she taught. After the first year, she received a regular salary. She was to spend the last eight years of her life teaching at this institution.

In the mid-1880s, Kovalevskaya made a second mathematical discovery of profound importance. Mathematical physics is made complicated by the fact that the differential equations used to describe even simple, idealized cases of physical laws are extremely difficult to solve. The obstacle consists of two parts. First, the equations must be reduced to a set of integrals to be evaluated; second, those integrals must be computed. In many important cases, such as the equations of the three-body problem, the first is impossible using only algebraic methods. When it is possible, the second is often impossible if only elementary functions are to be used. For example, the equation of pendulum motion can be reduced to an integral, but that integral involves the square root of a cubic or quartic polynomial; it is known as an *elliptic integral*. Such is the case in the phenomenon studied by Kovalevskaya, the motion of a rigid body about a fixed point.

The six equations of motion for a rigid body in general cannot be reduced to integrals using only algebraic transformations. In Kovalevskaya's day only two special cases were known in which such a reduction was possible, and the integrals in both cases were elliptic integrals. Only in the case of bodies satisfying the hypotheses of both of these cases simultaneously were the integrals elementary. With Weierstrass, however, Kovalevskaya had studied not merely elliptic integrals, but integrals of completely arbitrary algebraic functions. Such integrals were known as *abelian integrals* after Abel, the first person to make significant progress in studying them. She was not daunted by the prospect of working with such integrals, since she knew that the secret of taming them was to use the functions known as *theta functions*, which had been introduced earlier by Abel and his rival in the creation of elliptic function theory, Jacobi. All she had to do was reduce the equations of motion to integrals; evaluating them was within her power. Unfortunately, it turns out that the completely general set of such equations cannot be reduced to integrals. But Kovalevskaya found a new case, much less symmetric than the cases already known (due to Euler and Lagrange), in which this reduction was possible. The algebraic changes of variable by which she made this reduction are quite impressive, spread over some 16 pages of one of the papers she eventually published on this subject. Still more impressive is the 80-page argument that follows to evaluate these integrals, which turn out to be hyperelliptic, involving the square

¹It is now the University of Stockholm.

root of a fifth-degree polynomial. This work so impressed the leading mathematicians of Paris that they decided the time had come to propose a contest for work in this area. When the contest was held in 1888, Kovalevskaya submitted a paper and was awarded the prize. She had finally reached the top of her profession and was rewarded with a tenured position in Stockholm. Sadly, she was not to be in that lofty position for long. In January 1891 she contracted pneumonia while returning to Stockholm from a winter vacation in Italy and died on February 10.

35.1.1. Resistance from Conservatives

Lest it be thought that the existence of such a powerful talent as Sof'ya Kovalevskaya removed all doubt as to women's ability to create mathematics, we must point out that minds did not simply change immediately. Confronted with the evidence that good women mathematicians had already existed, the geometer Gino Loria (1862–1954) rationalized his continuing opposition to the admission of women to universities as follows, in an article in *Revue scientifique* in 1904:

As for... Sonja Kowalevsky, the collaboration [she] obtained from first-rate mathematicians prevents us from fixing with precision her mathematical role. Nevertheless what we know allows us to put the finishing touches on a character portrait of any woman mathematician. She is always a child prodigy, who, because of her unusual aptitudes, is admired, encouraged, and strongly aided by her friends and teachers. In childhood she manages to surpass her male fellow-students; in her youth she succeeds only in equalling them; while at the end of her studies, when her comrades of the other sex are progressing vigorously and boldly, she always seeks the support of a teacher, friend, or relative; and after a few years, exhausted by efforts beyond her strength, she finally abandons a work which is bringing her no joy.

Loria could have known better. Six years before Loria wrote these words Felix Klein (1849–1925) was quoted by the journal *Le progrès de l'est* as saying that he found his women students to be in every respect the equals of their male colleagues.

35.2. GRACE CHISHOLM YOUNG

Klein began taking on women students in the 1890s. The first of these students was Grace Chisholm, who completed the doctorate under his supervision in 1895 with a dissertation on the algebraic groups of spherical trigonometry. Her life and career were documented by her daughter and written up in an article by I. Grattan-Guinness (1972), which forms the basis for the present essay.

She was born on March 15, 1868, near London, the fifth child of parents of modest but comfortable means and the third child to survive. As a child she was stricken with polio and never completely recovered the use of her right hand. She was tutored at home and passed the Cambridge Senior Examination in 1885. She attended Girton College and met the prominent algebraist Arthur Cayley (1821–1895). Her impressions of him were not flattering. To her he seemed to be a lumbering intellectual dinosaur, preventing any new life

from emerging to enjoy the mathematical sunshine. In a colorful phrase, she wrote, “Cayley, unconscious himself of the effect he was having on his entourage, sat, like a figure of Buddha on its pedestal, dead-weight on the mathematical school of Cambridge” (Grattan-Guinness, 1972, p. 115).

In her first year at Cambridge, she might have been tutored by William Young (1863–1942), who later became her husband, except that she had heard that his teaching methods were ill-suited to young women. She found that Newnham College, the other women’s college at Cambridge, had a much more serious professional atmosphere than Girton. She made contacts there with two other young women who had the same tutor that she had. With the support of this tutor and her fellow women students, she began to move among the serious mathematicians at Cambridge and prepare to take the Tripos Examination.² In particular, she made friends with a student named Isabel Maddison (1869–1950) of Newnham College, who was being tutored by William Young. In 1890, after reading a few names of the top Wranglers, the moderator—W. W. Rouse Ball (1850–1925), the author of a best-selling popular history of mathematics—made a long pause to get the attention of the audience, then said in a loud, clear voice, “*Above* the Senior Wrangler: Fawcett, Newnham.” The young woman, Philippa Fawcett³ of Newnham College, had scored a major triumph for women’s education, being the top mathematics student at Cambridge in her year. No better role model can be imagined for students such as Isabel Maddison and Grace Chisholm. They finished first and second, respectively, in the year-end examinations at Girton College the following year. That fall, due to the absence of her regular tutor, Chisholm was forced to take lessons from William Young. In 1892 she ranked between the 23rd and 24th men on the Tripos, and Isabel Maddison finished in a tie with the 27th. (The rankings went as far as 112.) As a result, each received a First in mathematics. That same year they became the first women to attempt the Final Honours examinations at Oxford, where Chisholm obtained a First and Maddison a Second. This achievement made Chisholm the first person to obtain a First in any subject from both Oxford and Cambridge.⁴

Unfortunately, Cambridge did not offer Grace Chisholm support for graduate study, and her application to Cornell University in the United States was rejected. As an interesting irony, then, she was forced to apply to a university with a higher standard of quality than Cornell at the time, and one that was the mathematical equal of Cambridge: the University of Göttingen. There, thanks to the liberal views of Felix Klein and Friedrich Althoff,⁵ she was accepted, along with two young American women, Mary Frances (“May”) Winston (1869–1959) and Margaret Eliza Maltby (1860–1944). In 1895, Chisholm broached the subject of

²The Tripos Examination was a venerable tradition at Cambridge, dating back to Medieval times. A high-quality performance merited a First degree, lower quality a Second. Those who gained a First were called Wranglers. With modifications, the system continues at the present time.

³Philippa Garrett Fawcett (1868–1948) was the daughter of a professor of political economy at Cambridge. Her mother was a prominent advocate of women’s rights, and her sister was the first woman to obtain a medical degree at the University of St Andrews in Scotland. Philippa used her Cambridge education to go to the Transvaal in 1902 and help set up an educational system there. From 1905 to 1934 she was Director of Education of the London County Council.

⁴Isabel Maddison was awarded the Bachelor of Science degree at the University of London in 1892. She received the Ph.D. at Bryn Mawr in 1896 under the supervision of Charlotte Angus Scott (1858–1931, another alumna of Girton College and a student of Cayley). She taught at Bryn Mawr until her retirement in 1926.

⁵Althoff (1839–1908) was the Prussian Under-Secretary of Education and Cultural Affairs during the time of Kaiser Wilhelm II.

a Ph.D. with Klein, who agreed to use his influence in the faculty to obtain authorization for the degree. It turned out to be necessary to go all the way to the Ministry of Culture in Berlin and obtain permission from Althoff personally. Fortunately, Althoff continued to be an enthusiastic supporter, and her final oral examination took place on April 26 of that year. She passed it and was granted the Ph.D. *magna cum laude*. She herself could hardly take in the magnitude of her achievement. More than two decades had passed since the university had awarded the Ph.D. to Sof'ya Kovalevskaya *in absentia*. Grace Chisholm had become the first woman to obtain that degree in mathematics through regular channels anywhere in Germany. She and Mary Winston were left alone together for a few minutes, which they used "to execute a war dance of triumph." Her two companions Mary Winston and Margaret Maltby also received the Ph.D. degree at Göttingen, Maltby (in physics) in 1895 and Winston in 1896.⁶

Grace Chisholm sent a copy of her dissertation to her former tutor William Young, and in the fall of 1895 they began collaboration on a book on astronomy, a project that both soon forgot in the pleasant fog of courtship. They were married in June 1896. They planned a life in which Grace would do mathematical research and William would support the family by his teaching. Grace sent off her first research paper for publication, and William, who was then 33 years old, continued tutoring. Circumstances intervened, however, to change these plans. Cambridge began to reduce the importance of coaching, and the first of their four children was born in June 1897. Because of what they regarded as the intellectual dryness of Cambridge and the need for a more substantial career for William, they moved back to Germany in the autumn of 1897. With the help of Felix Klein, William sent off his first research paper to the London Mathematical Society. It was Klein's advice a few years later that caused both Youngs to begin working in set theory. William, once started in mathematics, proved to be a prolific writer. In the words of Grattan-Guinness (1972, p. 142), he "definitely belongs to the category of creative men who published more than was good for him." Moreover, he received a great deal of collaboration from his wife that, apparently by mutual consent, was not publicly acknowledged. He himself admitted that much of his role was to lay out for Grace problems that he couldn't solve himself. To the modern eye he appears too eager to interpret this situation by saying that "we are rising *together* to new heights." As he wrote to her:

The fact is that our papers ought to be published under our joint names, but if this were done neither of us get the benefit of it. No. Mine the laurels now and the knowledge. Yours the knowledge only. Everything under my name now, and later when the loaves and fishes are no more procurable in that way, everything or much under your name. [Grattan-Guinness, 1972, p. 141]

Perhaps the criticism Loria made of Sof'ya Kovalevskaya for obtaining help from first-rate mathematicians might more properly have been leveled against William Young. The rationalization in this quotation seems self-serving. Yet, the only person who could make

⁶Margaret Maltby taught at Barnard College (now part of Columbia University in New York) for 31 years and was chair of physics for 20 of those years. Mary Winston had studied at Bryn Mawr with Charlotte Angas Scott. She had met Felix Klein at the World's Columbian Exposition in Chicago in 1893 and had moved to Göttingen at his invitation. After returning to the United States she taught at Kansas State Agricultural College, married Henry Newson, a professor of mathematics at the University of Kansas, bore three children, and went back to teaching after Henry's early death. From 1921 to 1942 she taught at Eureka College in Illinois.

that judgment, Grace Chisholm Young herself, never gave any hint that she felt exploited, and William was certainly a very talented mathematician in his own right, whose talent manifested itself rather late in life. And one cannot deny that, given the state of society at the time, the situation William Young is describing was very likely the best option for both of them.

In 1903 Cambridge University Press agreed to publish a work on set theory under both their names. That book appeared in 1906; a book on geometry appeared under both names in 1905. Grace was busy bearing children all this time (their last three children were born in 1903, 1904, and 1908) and studying medicine. She began to write mathematical papers under her own name in 1913, after William took a position in Calcutta, which of course required him to be away for long periods of time. These papers, especially her paper on the differentiability properties of completely arbitrary functions, added to her reputation and were cited in textbooks on measure theory for many decades.

Sadly, the fanaticism of World War I caused some strains between the Youngs and their old mentor Felix Klein. As a patriotic German, Klein had signed a declaration of support for the German position at the beginning of the war. Four years later, as the defeat of Germany drew near, Grace wrote to him, asking him to withdraw his signature. Of course, propaganda had been intense in all the belligerent countries during the war, and even the mildest-mannered people tended to believe what they were told and to hate the enemy. Klein replied diplomatically, saying that, "Everyone will hold to his own country in light and dark days, but we must free ourselves from passion if international cooperation such as we all desire is to assert itself again for the good of the whole" (Grattan-Guinness, 1972, p. 160). If only other scholars, in other countries, had been as magnanimous as Klein, German scholars might have had less justification for complaining of exclusion in the bitter postwar period. At least there was no irreparable breach between the Youngs and Klein. When Klein died in 1925, his widow thanked the Youngs for sending their sympathy, saying, "From all over the world I received such lovely letters full of affection and gratitude, so many tell me that he showed them the way on which their life was built. I had him for fifty years, this wonderful man; how privileged I am above most women. . ." (Grattan-Guinness, 1972, p. 171).

All four of their children eventually obtained doctoral degrees, and the pair had good grounds for being well-satisfied with their married life. When World War II began in September 1939, they were on holiday in Switzerland, and there was fear that Switzerland would be invaded. Grace immediately returned to England, but William stayed behind. The fall of France in 1940 enforced a long separation on them. The health of William, who was by then in his late 70s, declined rapidly, and he died in a nursing home in June 1942. Grace survived for nearly two more years, dying in March 1944. Grattan-Guinness (1972, p. 181) has eloquently characterized this remarkable woman:

She knew more than half a dozen languages herself, and in addition she was a good mathematician, a virtually qualified medical doctor, and in her spare time, pianist, poet, painter, author, Platonic and Elizabethan scholar—and a devoted mother to all her children. And in the blend of her rôles as scholar and as mother lay the fulfillment of her complicated personality.

35.3. EMMY NOETHER

Sof'ya Kovalevskaya and Grace Chisholm Young had had to improvise their careers, taking advantage of the opportunities that arose from time to time. One might have thought that

Amalie Emmy Noether was better situated in regard to both the number of opportunities arising and the ability to take advantage of them. After all, she came a full generation later than Kovalevskaya, the University of Göttingen had been awarding degrees to women for five years when she enrolled, and she was the eldest child of the distinguished mathematician Max Noether. According to Dick (1981), on whose biography of her the following account is based, she was born on March 23, 1882 in Erlangen, Germany, where her father was a professor of mathematics. She was to acquire three younger brothers in 1883, 1884, and 1889. Her childhood was quite a normal one for a girl of her day, and at the age of 18 she took the examinations for teachers of French and English, scoring very well. This achievement made her eligible to teach modern languages at women's educational institutions. She decided instead to attend the University of Erlangen. There, she was one of only two women in the student body of 986, and she was only an auditor, preparing simultaneously to take the graduation examinations in Nürnberg. After passing these examinations, she went to the University of Göttingen for one year, again not as a matriculated student. If it seems strange that Grace Chisholm was allowed to matriculate at Göttingen and Emmy Noether was not, the explanation seems to be precisely that Emmy Noether was a German.

In 1904 she was allowed to matriculate at Erlangen, where she wrote a dissertation under the direction of Paul Gordan (1837–1912). Gordan was a constructivist and disliked abstract proofs. According to Kowalewski (1950, p. 25) he is said to have remarked of one proof of the Hilbert basis theorem, "That is no longer mathematics; that is theology." In her dissertation, Emmy Noether followed Gordan's constructivist methods; but she was later to become famous for work done from a much more abstract point of view. She received the doctorate *summa cum laude* in 1907. Thus, she overcame the first two obstacles to a career in mathematics with only a small amount of difficulty, not much more than faced by her brother Fritz (1884–1941), who was also a mathematician. That third obstacle, however, finding work at a university, was formidable. Emmy Noether spent many years working without salary at the Mathematical Institute in Erlangen. This position enabled her to look after her father, who had been frail since he contracted polio at the age of 14. It also allowed her to continue working on mathematical ideas. For nearly two decades she corresponded with Gordan's successor in Erlangen, Ernst Fischer (1875–1954), who is best remembered for having discovered the Riesz–Fischer theorem independently of F. Riesz (1880–1956). By staying in touch with the mathematical community and giving lectures on her discoveries, she kept her name before certain influential mathematicians, namely David Hilbert (1862–1943) and Felix Klein,⁷ and in 1915 she was invited to work as a *Privatdozent* in Göttingen, the same rank originally offered to Kovalevskaya at Stockholm in 1883. Over the next four years Klein and Hilbert used all their influence to get her a regular appointment at Göttingen; during part of that time she lectured for Hilbert in mathematical physics. That work led her to a theorem in general relativity that was highly praised by both Hilbert and Einstein. Despite this brilliant work, however, she was not allowed to pass the *Habilitation* needed to acquire a professorship. Only after the German defeat in World War I, which was followed by the abdication of the Kaiser and a general spirit of reform in Germany, was she allowed to "habilitate." Between Sof'ya Kovalevskaya and Emmy Noether there was a curious kind of symmetry: Kovalevskaya was probably aided in her efforts to become a student in Berlin because many of the students were away at war at the time. Noether was

⁷Klein wrote to Hilbert, "You know that Fräulein Noether is continually advising me in my projects and that it is really through her that I have become competent in the subject." (Dick, 1981, p. 31)

aided in her efforts to become a professor by an influx of returning war veterans. She began lecturing in courses offered under the name Dr. Emmy Noether (without any mention of Hilbert) in the fall of 1919. Through the efforts of Richard Courant (1888–1972) she was eventually granted a small salary for her lectures.

In the 1920s she moved into the area of abstract algebra, and it is in this area that mathematicians know her work best. Noetherian rings became a basic area of study after her work, which became part of a standard textbook by her student Bartel Leendert van der Waerden (1903–1996). He later described her influence on this work (1975, p. 32):

When I came to Göttingen in 1924, a new world opened up before me. I learned from Emmy Noether that the tools by which my questions could be handled had already been developed by Dedekind [Richard Dedekind (1831–1916) and Weber [Heinrich Weber, 1842–1913)], by Hilbert, Lasker [Emanuel Lasker (1868–1941)] and Macaulay [Francis Sowerby Macaulay (1862–1937)], by Steinitz [Ernst Steinitz (1871–1928)] and by Emmy Noether herself.

Of all the women we have discussed Emmy Noether was unquestionably the most talented mathematically. Her work, both in quantity and quality, places her in the elite of twentieth-century mathematicians, and it was recognized as such during her lifetime. She became an editor of *Mathematische Annalen*, one of the two or three most prestigious journals in the world. She was invited to speak at the International Congress of Mathematicians in Bologna in 1928 and in Zürich in 1932, when she shared with Emil Artin (1898–1962) a prestigious prize for the advancement of mathematical knowledge. This recognition was clear and simple proof of her ability. Hilbert's successor in Göttingen, Hermann Weyl (1885–1955), made this point when wrote her obituary:

When I was called permanently to Göttingen in 1930, I earnestly tried to obtain from the Ministerium a better position for her, because I was ashamed to occupy such a preferred position beside her, whom I knew to be my superior as a mathematician in many respects. I did not succeed, nor did an attempt to push through her election as a member of the Göttinger Gesellschaft der Wissenschaften. Tradition, prejudice, external considerations, weighted the balance against her scientific merits and scientific greatness, by that time denied by no one. In my Göttingen years, 1930–1933, she was without doubt the strongest center of mathematical activity there. [Dick, 1981, p. 169]

To have been recognized by one of the twentieth century's greatest mathematicians as “the strongest center of mathematical activity” at a university that was second to none in the quality of its research is high praise indeed. It is unfortunate that this recognition was beyond the capability of the Ministerium. The year 1932 was to be the summit of Noether's career. The following year, the advanced culture of Germany, which had enabled her to develop her talents to their fullest, turned its back on its own brilliant past and plunged into the nightmare of Nazism. Despite extraordinary efforts by the greatest scientists on her behalf, Noether was removed from the position that she had achieved through such a long struggle and the assistance of great mathematicians. Along with hundreds of other Jewish mathematicians, including her friends Richard Courant and Hermann Weyl (who was not Jewish, but whose wife was), she had to find a new life in a different land. She accepted a visiting professorship at Bryn Mawr, which allowed her also to lecture at the Institute

for Advanced Study in Princeton.⁸ Despite the gathering clouds in Germany, she returned there in 1934 to visit her brother Fritz, who was about to seek asylum in the Soviet Union. (Ironically, he was arrested in 1937, during one of the many purges conducted by Stalin, and executed as a German spy on the day the Germans occupied Smolensk in 1941.) She returned to Bryn Mawr in the spring of 1934.

Weyl, who went to Princeton in 1933, expressed his indignation at the Nazi policy of excluding “non-Aryans” from teaching. In a letter sent to Heinrich Brandt (1886–1954) in Halle he gave his opinion of Nazi-sympathizing intellectuals like Oswald Spengler and Ludwig Bieberbach⁹:

What impresses me most about Emmy Noether is that her research has become more and more concrete and profound. Why should this Jewess not work in the area that has led to such great achievements in the hands of the “Aryan” Dedekind? I am happy to leave it to Herrn Spengler and Bieberbach to assign mathematical modes of thought according to cultures and races. [Jentsch, 1986, p. 9]

At Bryn Mawr she was a great success and an inspiration to the women studying there. She taught several graduate and postdoctoral students who went on to successful careers, including her former assistant from Göttingen, Olga Taussky (1906–1995), who was forced to leave a tutoring position in Vienna in 1933. Her time, however, was to be very brief. She developed a tumor in 1935, but she does not seem to have been worried about its possible consequences. It was therefore a great shock to her colleagues in April 1935 when, after an operation at Bryn Mawr Hospital that seemed to offer a good prognosis, she developed complications and died within a few hours.

QUESTIONS

Historical Questions

- 35.1. For what mathematical achievements is Sof’ya Kovalevskaya best remembered?
- 35.2. What events turned Grace Chisholm Young toward mathematics, and how was she able to fulfill her ambition to become a mathematician?
- 35.3. What special contribution did Bryn Mawr College make toward the mathematical education of women?
- 35.4. In what areas of mathematics was Emmy Noether a world leader in research?

⁸There was no chance of her lecturing at Princeton University itself, which was all-male at the time.

⁹Oswald Spengler (1880–1936) was a German philosopher of history, best known for having written *Der Untergang des Abendlandes (The Decline of the West)*. His philosophy of history, which Weyl alludes to in this quote, suited the Nazis. Although at first sympathetic to them, he was repelled by their crudity and their antisemitism. By the time Weyl wrote this letter, the Nazis had banned all mention of Spengler on German radio. Ludwig Bieberbach (1886–1982) was a mathematician of some talent who worked in Berlin during the Nazi era and edited the Party-approved journal *Deutsche Mathematik*. At the time when Weyl wrote this letter, Bieberbach was wearing a Nazi uniform to the university and enthusiastically endorsing the persecution of non-Aryans.

Questions for Reflection

- 35.5.** What were the advantages and disadvantages of marriage for a woman seeking an academic career before the twentieth century? How much of this depended on the particular choice of a husband at each stage of the career? The cases of Sof'ya Kovalevskaya, Grace Chisholm Young, and Emmy Noether will be illuminating, but it will be useful to seek more detailed sources than the narratives above.
- 35.6.** How important is (or was) encouragement from family and friends in the decision to study science? How important is it to have a mentor, an established professional in the same field, to help orient early career decisions? How important is it for a young woman to have an older woman as a role model? Try to answer these questions along a scale from “not at all important” through “somewhat important” and “very important” to “essential.” Use the examples of the women whose careers are sketched above to support your rankings.
- 35.7.** How strong are the claims that Loria adduces in his argument against admitting women to universities? Were all the women discussed here encouraged by their families when they were young? Is it really true that it is impossible to “fix with precision” the original contributions of Sof'ya Kovalevskaya? Would collaboration with other mathematicians make it impossible to “fix with precision” the work of any male mathematicians? Consider also the case of the three women discussed above. Is it true that they were exhausted after finishing their education?

Next, consider that universities select the top students in high school classes for admission, so that a student who excelled the other students in high school might be able at best to equal the other students at a university. Further selections for graduate school, then for hiring at universities of various levels of prestige, then for academic honors, provide layer after layer of filtering. Except for an extremely tiny elite, those who were at the top at one stage find themselves in the middle at the next and eventually reach (what is ideally) a level commensurate with their talent. What conclusions could be justified in regard to any gender link in this universal process, based on a sample of fewer than five women? And how can Loria be sure he knows their proper level when all the women up to the time of writing were systematically locked out of the best opportunities for professional advancement? Look at the twentieth century and see what becomes of Loria's argument that women never reach the top.

Finally, examine Loria's argument in the light of the cold facts of society: A woman who wished to have a career in mathematics would naturally be well advised to find a mentor with a well-established reputation, as Sof'ya Kovalevskaya did. A woman who did not do that would have no chance of being cited by Loria as an example, since she would never have been heard of. Is this argument not a classical example of Catch-22?

- 35.8.** The primary undergraduate competition for mathematics majors in the United States is the Putnam Examination, administered the first weekend in December each year by the Mathematical Association of America. In addition to its rankings for the top teams and the top individuals, this examination also provides, for women who choose to enter, a prize for the highest-ranking woman. (The people grading the examinations do not know the identities of the entrants, and a woman can enter this

competition without giving her name or the name of her university to the graders.) Is this policy an important affirmative-action step to encourage talented young women in mathematical careers, or does it “send the wrong message,” implying that women cannot compete with men on an equal basis in mathematics? If you consider it a good thing, how long should it be continued? Forever? If not, what criterion should be used to determine when to discontinue the separate category?

Probability

One important part of modern mathematics that has not yet been mentioned is the theory of probability. Besides being a mathematical subject with its own special principles, it provides the mathematical apparatus for another discipline (statistics), which has perhaps more applications in the modern world than all of mathematics and also its own theoretical side. Unfortunately, we do not have space to discuss more than a few incidents in the history of statistics.

The word *probability* is related to the English words *probe*, *probation*, *prove*, and *approve*. All of these words originally had a sense of *testing* or *experimenting*,¹ reflecting their descent from the Latin *probo*, which has these meanings. In other languages the word used in this mathematical sense has a meaning more like *plausibility*,² as in the German *Wahrscheinlichkeit* (literally, *truth resemblance*) or the Russian *veroyatnost'* (literally, *credibility*, from the root *ver-*, meaning *faith*). The concept is very difficult to define in declarative sentences, precisely because it refers to phenomena that are normally described in the subjunctive mood. This mood has nearly disappeared in modern English; it clings to a precarious existence in the past tense, “If it were true that . . .” having replaced the older “If it be true that . . .”. The language of Aristotle and Plato, however, who were among the first people to discuss chance philosophically, had two such moods, the subjunctive and the optative, by which it was possible to express the difference between what *would* happen and what *might* happen. As a result, they could express more easily than we the intuitive concepts involved in discussing events that are imagined rather than observed.

Intuitively, probability attempts to express the relative strength of the feeling of confidence we have that an event will occur. How surprised would we be if the event happened? How surprised would we be if it did not happen? Because we do have different degrees of confidence in certain future events, quantitative concepts become applicable to the study of probability. Generally speaking, if an event occurs essentially all the time under specified conditions, such as an eclipse of the sun, we use a deterministic model (geometric astronomy, in this case) to study and predict it. If it occurs sometimes under conditions frequently associated with it, we rely on probabilistic models. Some earlier scientists and philosophers

¹The common phrase “the exception that proves the rule” is nowadays misunderstood and misused because of this shift in the meaning of the word *prove*. Exceptions *test* rules, they do not *prove* them in the current sense of that word. In fact, quite to the contrary, exceptions *disprove* rules.

²Here is another interesting word etymology. The root is *plaudo*, meaning *strike*, but specifically meaning to clap one’s hands together, to applaud. Once again, *approval* is involved in the notion of probability.

regarded probability as a measure of our ignorance. Kepler, for example, believed that the supernova of 1604 in the constellation Serpent may have been caused by a random collision of particles; but in general he was a determinist who thought that our uncertainty about a roll of dice was merely a matter of insufficient data being available. He admitted, however, that he could find no law to explain the apparently random pattern of eccentricities in the elliptical orbits of the six planets known to him.

Once the mathematical subject got started, it developed a life of its own, in which theorems could be proved with the same rigor as in any other part of mathematics. Only the application of those theorems to the physical world remained and remains clouded by doubt. We use probability informally every day, as the weather forecast informs us that the chance of rain is 30% or 80% or 100%,³ or when we are told that one person in 30 will be afflicted with Alzheimer's disease between the ages of 65 and 74. Much of the public use of such probabilistic notions is, although not meaningless, at least of questionable value. For example, we are told that the life expectancy of an average American is now 77 years. Leaving aside the many assumptions of environmental and political stability used in the model that produced this number, we should at least ask one question: Can the number be related to the life of any person in any meaningful way? What plans can one base on it, since anyone may die on any given day, yet very few people can confidently rule out the possibility of living past age 90?⁴

The many uncertainties of everyday life, such as the weather and our health, occur mixed with so many possibly relevant variables that it would be difficult to distill a theory of probability from those intensely practical matters. What is needed is a simpler and more abstract model from which principles can be extracted and gradually made more sophisticated. The most obvious and accessible such models are games of chance. On them, probability can be given a quantitative and empirical formulation, based on the frequency of wins and losses. At the same time, the imagination can arrange the possible outcomes symmetrically and in many cases assign equal probabilities to different events. Finally, since money generally changes hands at the outcome of a game, the notion of a random variable (payoff to a given player, in this case) as a quantity assuming different values with different probabilities can be modeled.

36.1. CARDANO

The mathematization of probability began in sixteenth-century Italy with Cardano. Todhunter (1865), on whose work the following discussion of Cardano's book is based, reports (p. 3) that Cardano gave the following table of values for a roll of three dice.

1	2	3	4	5	6	7	8	9	10	11	12
108	111	115	120	126	133	33	36	37	36	33	26

³These numbers are generated by computer models of weather patterns for squares in a grid representing a geographical area. The modeling of their accuracy also uses probabilistic notions.

⁴The Russian mathematician Yu. V. Chaikovskii (2001) believes that some of this cloudiness is about to be removed with the creation of a new science he calls *aleatics* (from the Latin word *alea*, meaning *dice-play* or *gambling*). We must wait and see. A century ago, other Russian mathematicians confidently predicted a bright future for *arithmology*.

This table is enigmatic. Since it is impossible to roll a 1 with three dice, the first entry should perhaps be interpreted as the number of ways in which 1 may appear on *at least* one of the three dice. If so, then Cardan has got it wrong. One can imagine him thinking that if a 1 appears on one of the dice, the other two may show 36 different numbers, and since there are three dice on which the 1 may appear, the total number of ways of rolling a 1 must be $3 \cdot 36$ or 108. That way of counting ignores the fact that in some of these cases 1 appears on two of the dice or all three. By what is now known as the inclusion-exclusion principle, the total should be $3 \cdot 36 - 3 \cdot 6 + 1 = 91$. But it is difficult to say what Cardano had in mind. The number 111 given for 2 may be the result of the same count, increased by the three ways of choosing two of the dice to show a 1. Todhunter worked out a simple formula giving these numbers, but could not imagine any gaming rules that would correspond to them. If indeed Cardano made mistakes in his computations, he was not the only great mathematician to do so.

Cardano's *Liber de ludo* (*Book on Gambling*) was published about a century after his death. In this book Cardano introduces the idea of assigning a probability p between 0 and 1 to an event whose outcome is not certain. The principal applications of this notion were in games of chance, where one might bet, for example, that a player could roll a 6 with one die given three chances. The subject is not developed in detail in Cardano's book, much of which is occupied by descriptions of the actual games played. Cardano does, however, state the multiplicative rule for a run of successes in independent trials. Thus the probability of getting a six on each of three successive rolls with one die is $(\frac{1}{6})^3$. Most important, he recognized the real-world application of what we call the law of large numbers, saying that when the probability for an event is p , then after a large number n of repetitions, the number of times it will occur does not lie far from the value np . This law says that it is not certain that the number of occurrences will be near np , but "that is where the smart money bets."

36.2. FERMAT AND PASCAL

After a bet has been made and before it is settled, a player cannot unilaterally withdraw from the bet and recover her or his stake. On the other hand, an accountant computing the net worth of one of the players ought to count part of the stake as an asset owned by that player; and perhaps the player would like the right to sell out and leave the game. What would be a fair price to charge someone for taking over the player's position? More generally, what happens if the game is interrupted? How are the stakes to be divided? The principle that seemed fair was that, *regardless of the relative amount of the stake each player had bet, at each moment in the game a player should be considered as owning the portion of the stakes equal to that player's probability of winning at that moment*. Thus, the net worth of each player is constantly changing as the game progresses, in accordance with what we now call *conditional probability*. Computing these probabilities in games of chance usually involves the combinatorial counting techniques the reader may have encountered in elementary discussions of probability. Problems of this kind were discussed in correspondence between Pascal and Fermat in the mid-seventeenth century.

A French nobleman, the Chevalier de Méré, who was fond of gambling, proposed to Pascal the problem of dividing the stakes in a game where one player has bet that a six will appear in eight rolls of a single die, but the game is terminated after three unsuccessful tries. Pascal wrote to Fermat that the player should be allowed to sell the throws one at a time. If the first throw is foregone, the player should take one-sixth of the stake, leaving five-sixths.

Then if the second throw is also foregone, the player should take one-sixth of the remaining five-sixths or $\frac{5}{36}$, and so on. In this way, Pascal argued that the fourth through eighth throws were worth $\frac{1}{6}[(\frac{5}{6})^3 + (\frac{5}{6})^4 + (\frac{5}{6})^5 + (\frac{5}{6})^6 + (\frac{5}{6})^7]$.

This expression is the value of those throws *before* any throws have been made. If, after the bets are made but before any throws of the die have been made, the bet is changed and the players agree that only three throws shall be made, then the player holding the die should take this portion of the stakes as compensation for sacrificing the last five throws. Remember, however, that the net worth of a player is constantly changing as the game progresses and the probability of winning changes. The value of the fourth throw, for example, is smaller to begin with, since there is some chance that the player will win before it arrives, in which case it will not arrive. At the beginning of the game, the chance of winning on the fourth roll is $(\frac{5}{6})^3 \frac{1}{6}$. Here the factor $(\frac{5}{6})^k$ in each term represents the probability that the player *will not have won* in the first k terms. After three unsuccessful throws, however, the probability that the player “will not have” won (that is to say, *did not win*) on the first three throws is 1, and so the probability of winning on the fourth throw becomes $\frac{1}{6}$.

Fermat expressed the matter as follows:

[T]he three first throws having gained nothing for the player who holds the die, the total sum thus remaining at stake, he who holds the die and who agrees not to play his fourth throw should take $\frac{1}{6}$ as his reward. And if he has played four throws without finding the desired point and if they agree that he shall not play the fifth time, he will, nevertheless, have $\frac{1}{6}$ of the total for his share. Since the whole sum stays in play it not only follows from the theory, but it is indeed common sense that each throw should be of equal value.

Pascal wrote back to Fermat, proclaiming himself satisfied with Fermat’s analysis and overjoyed to find that “the truth is the same at Toulouse and at Paris.”

36.3. HUYGENS

The Dutch mathematical physicist Christiaan Huygens (1629–1695), author of a very influential book on optics, wrote a treatise on probability in 1657. His *De ratiociniis in ludo aleae* (*On Reasoning in a Dice Game*) consisted of 14 propositions and contained some of the results of Fermat and Pascal. In addition, Huygens considered multinomial problems, involving three or more players. Cardano’s idea of an *estimate of the expectation* was elaborated by Huygens. He asserted, for example, that if there are p (equally likely) ways for a player to gain a and q ways to gain b , then the player’s expectation is $(pa + qb)/(p + q)$.

Even simple problems involving these notions can be subtle. For example, Huygens considered two players A and B taking turns rolling a pair of dice, with A going first. Any time A rolls a 6, A wins; any time B rolls a 7, B wins. What are the relative chances of winning? (The answer to that question would determine the fair proportions of the stakes to be borne by the two players.) Huygens concluded (correctly) that the odds were 31:30 in favor of B , that is, A ’s probability of winning was $\frac{30}{61}$ and B ’s probability was $\frac{31}{61}$.

36.4. LEIBNIZ

Although Leibniz wrote a full treatise on combinatorics, which provides the mathematical apparatus for computing many probabilities in games of chance, he did not himself gamble.

But he did analyze many games of chance and suggest modifications of them that would make them fair (zero-sum) games. Some of his manuscripts on this topic have been analyzed by De Mora-Charles (1992). One of the games he analyzed is known as quinquenove. This game is played between two players using a pair of dice. One of the players, called the banker, rolls the dice, winning if the result is either a double or a total number of spots showing equal to 3 or 11. There are thus 10 equally likely ways for the banker to win with this roll, out of 36 equally likely outcomes. If the banker rolls a 5 or 9 (hence the name “quinquenove”), the other player wins. The other player has eight ways of winning of the equally likely 36 outcomes, leaving 18 ways for the game to end in a draw. The reader may be amused to learn that the great mathematician Leibniz, author of *De arte combinatoria*, confused permutations and combinations in his calculations for this game and got the probabilities wrong.

36.5. THE ARS CONJECTANDI OF JAMES BERNOULLI

One of the founding documents of probability theory was published in 1713, eight years after the death of its author, Leibniz’ disciple James Bernoulli. This work, *Ars conjectandi* (*The Art of Prediction*), moved probability theory beyond the limitations of analyzing games of chance. It was intended by its author to apply mathematical methods to the uncertainties of life. As he said in a letter to Leibniz, “I have now finished the major part of the book, but it still lacks the particular examples, the principles of the art of prediction that I teach how to apply to society, morals, and economics. . . .” That was an ambitious undertaking, and Bernoulli had not quite finished the work when he died in 1705.

Bernoulli noted an obvious gap between theory and application, saying that only in simple games such as dice could one apply the equal-likelihood approach of Huygens, Fermat and Pascal, whereas in the cases of practical importance, such as human health and longevity, no one had the power to construct a suitable model. He recommended statistical studies as the remedy to our ignorance, saying that if 200 people out of 300 of a given age and constitution were known to have died within 10 years, it was a 2-to-1 bet that any other person of that age and constitution would die within a decade.

In this treatise, Bernoulli reproduced the problems solved by Huygens and gave his own solution of them. He considered what are now called *Bernoulli trials*. These are repeated experiments in which a particular outcome either happens (success) with probability b/a or does not happen (failure) with probability c/a , the same probability each time the experiment is performed, each outcome being independent of all others. (A simple nontrivial example is rolling a single die, counting success as rolling a 3. Then the probabilities are $\frac{1}{6}$ and $\frac{5}{6}$.) Since $b/a + c/a = 1$, Bernoulli saw that the binomial expansion would be useful in computing the probability of getting at least m successes in n trials. He gave that probability as

$$\sum_{k=m}^n \binom{n}{k} \left(\frac{b}{a}\right)^k \left(\frac{c}{a}\right)^{n-k}.$$

It was, incidentally, in this treatise, when computing the sum of the r th powers of the first n integers, that Bernoulli introduced what are now called the *Bernoulli numbers*,⁵

⁵As mentioned in Chapter 24, a table of these numbers can be found in Seki Kōwa’s posthumously published *Katsuyō Sampō*, which appeared the year before Bernoulli’s work.

$1, \frac{1}{2}, A, B, \dots$ defined by the formula

$$\sum_{k=1}^n k^r = \frac{n^{r+1}}{r+1} + \frac{n^r}{2} + \frac{r}{2}An^{r-1} + \frac{r(r-1)(r-2)}{2 \cdot 3 \cdot 4}Bn^{r-3} + \dots$$

Nowadays we define these numbers as $B_0 = 1$, $B_1 = -\frac{1}{2}$, and then $B_2 = A = \frac{1}{6}$, $B_3 = 0$, $B_4 = B$, and so forth. He illustrated his formula by finding

$$\sum_{k=1}^{1000} k^{10} = 91409924241424243424241924242500.$$

36.5.1. The Law of Large Numbers

Bernoulli imagined an urn containing numbers of black and white pebbles, whose ratio is to be determined by sampling with replacement. It is possible that you will always get a white pebble, no matter how many times you sample. However, if black pebbles constitute a significant proportion of the contents of the urn, this outcome is very unlikely. After discussing the degree of certainty that would suffice for practical purposes (he called it *virtual certainty*),⁶ he noted that this degree of certainty could be attained empirically by taking a sufficiently large sample. The probability that the empirically determined ratio would be close to the true ratio increases as the sample size increases, but the result would be accurate only within certain limits of error, and

... we can attain any desired degree of probability that the ratio found by our many repeated observations will lie between these limits.

This last assertion is an informal statement of the law of large numbers for Bernoulli trials. If the probability of the outcome is p and the number of trials is n , this law can be phrased precisely by saying that for any $\varepsilon > 0$ there exists a number n_0 such that if m is the number of times the outcome occurs in n trials and $n > n_0$, the probability that the inequality $|(m/n) - p| > \varepsilon$ will hold is less than ε .⁷ Bernoulli stated this principle in terms of the segment of the binomial series of $(r + s)^{n(r+s)}$ consisting of the n terms on each side of the largest term (the term containing $r^{nr}s^{ns}$), and he proved it by giving an estimate on n sufficient to make the ratio of this sum to the sum of the remaining terms at least c , where c is specified in advance.

⁶This phrase is often translated more literally as *moral* certainty, which has the wrong connotation.

⁷Probabilists say that the frequency of successes converges "in probability" to the probability of success at each trial. Analysts say it converges "in measure." There is also a strong law of large numbers, more easily stated in terms of independent random variables, which asserts that (under suitable hypotheses) there is a set of probability 1 on which the convergence to the mean occurs. That is, the convergence is "almost surely," as probabilists say, and "almost everywhere," as analysts phrase the matter. On a finite measure space such as a probability space, almost everywhere convergence implies convergence in measure, but the converse is not true.

36.6. DE MOIVRE

In 1711, even before the appearance of James Bernoulli's treatise, another ground-breaking book on probability appeared, the *Doctrine of Chances*, written by Abraham De Moivre (1667–1754), a French Huguenot who took refuge in England after 1685, when Louis XIV revoked the Edict of Nantes, issued by Henri IV to put an end to religious war in France by guaranteeing the civil rights of the Huguenots upon his accession to the throne in 1598.⁸ De Moivre's book went through several editions. Its second edition, which appeared in 1738, introduced a significant piece of numerical analysis, useful for approximating sums of terms of a binomial expansion $(a + b)^n$ for large n . De Moivre had published the work earlier in a paper written in 1733. Having no notation for the base e , which was introduced by Euler a few years later, De Moivre simply referred to the hyperbolic (natural) logarithm and "the number whose logarithm is 1." De Moivre first considered only the middle term of the expansion. That is, for an even power $n = 2m$, he estimated the term

$$\binom{2m}{m} = \frac{(2m)!}{(m!)^2}$$

and found it equal to $\frac{2^{2m}}{B\sqrt{n}}$, where B was a constant for which he knew only an infinite series. At that point, he got stuck, as he admitted, until his friend James Stirling (1692–1770) showed him that "the Quantity B did denote the Square-root of the circumference of a circle whose Radius is Unity."⁹ In our terms, $B = \sqrt{2\pi}$, but De Moivre simply wrote c for B . Without having to know the exact value of B , De Moivre was able to show that "the Logarithm of the Ratio, which a Term distant from the middle by the Interval ℓ , has the the middle Term, is [approximately, for large n] $-\frac{2\ell\ell}{n}$." In modern language,

$$\binom{2n}{n+\ell} / \binom{2n}{n} \approx e^{-2\ell^2/n}.$$

De Moivre went on to say, "The Number, which answers to the Hyperbolic Logarithm $-2\ell\ell/n$, [is]

$$1 - \frac{2\ell\ell}{n} + \frac{4\ell^4}{2nn} - \frac{8\ell^6}{6n^3} + \frac{16\ell^8}{24n^4} - \frac{32\ell^{10}}{120n^5} + \frac{64\ell^{12}}{720n^6}, \text{ \&c.}''$$

By scaling, De Moivre was able to estimate segments of the binomial distribution. In particular, the fact that the numerator was ℓ^2 and the denominator n allowed him to estimate the probability that the number of successes in Bernoulli trials would be between fixed limits. He came close to noticing that the natural unit of probability for n trials was a multiple of \sqrt{n} . In 1893, this natural unit of measure for probability was named the *standard deviation* by

⁸The spirit of sectarianism has infected historians to the extent that Catholic and Protestant biographers of De Moivre do not agree on how long he was imprisoned in France for being a Protestant. They do agree that he was imprisoned, however. To be fair to the French, they did elect him a member of the Academy of Sciences a few months before his death.

⁹The approximation $n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$ is now called *Stirling's formula*.

the British mathematician Karl Pearson (1857–1936). For Bernoulli trials with probability of success p at each trial the standard deviation is $\sigma = \sqrt{np(1-p)}$.

For what we would call a coin-tossing experiment in which $p = \frac{1}{2}$ —he imagined tossing a metal disk painted white on one side and black on the other—de Moivre observed that with 3600 coin tosses, the odds would be more than 2 to 1 against a deviation of more than 30 “heads” from the expected number of 1800. The standard deviation for this experiment is exactly 30, and 68 percent of the area under a normal curve lies within one standard deviation of the mean. De Moivre could imagine the bell-shaped normal curve that we are familiar with, but he did not give it the equation it now has ($y = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$). Instead he described it as the curve whose ordinates were numbers having certain logarithms. What seems most advanced in his analysis is that he recognized the area under the curve as a probability and computed it by a mechanical quadrature method that he credited jointly to Newton, Roger Cotes, James Stirling, and himself. This tendency of the distribution density of the average of many independent, identically distributed random variables to look like the bell-shaped curve is called the *central limit theorem*.

36.7. THE PETERSBURG PARADOX

Soon after its introduction by Huygens and James Bernoulli the concept of mathematical expectation came in for some critical appraisal. While working in the Russian Academy of Sciences, Daniel Bernoulli discussed the problem now known as the *Petersburg paradox* with his brother Nicholas (1695–1726, known as Nicholas II). We can describe this paradox informally as follows. Suppose that you toss a coin until heads appears. If it appears on the first toss, you win \$2, if it first appears on the second toss, you win \$4, and so on; if heads first appears on the n th toss, you win 2^n dollars. How much money would you be willing to pay to play this game? Now by “rational” computations the expected winning is infinite, being $2 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4} + 8 \cdot \frac{1}{8} + \dots$, so that you should be willing to pay, say, \$10,000 to play each time. On the other hand, who would bet \$10,000 knowing that there was an even chance of winning back only \$2, and that the odds are 7 to 1 against winning more than \$10? Something more than mere expectation was involved here.

Daniel Bernoulli discussed the matter at length in an article in the *Commentarii* of the Petersburg Academy for 1730–1731 (published in 1738). He argued for the importance of something that we now call *utility*. If you already possess an amount of money x and you receive a small additional amount of money dx , how much *utility* does the additional money have for you, subjectively? Bernoulli assumed that the increment of utility dy was directly proportional to dx and inversely proportional to x , so that

$$dy = \frac{k dx}{x},$$

and as a result, the total utility of personal wealth is a logarithmic function of total wealth.

One consequence of this assumption is a law of diminishing returns: The additional satisfaction from additional wealth decreases as wealth increases. Bernoulli used this idea to explain why a rational person would refuse to play the game. Obviously, the expected gain in utility from each of these wins, being proportional to the logarithm of the money gained, has a finite total, and so one should be willing to pay only an amount of money that has an equal utility to the gambler. A different explanation, which seems to have been given

first by the mathematician John Venn (1834–1923) of Caius¹⁰ College, Cambridge in 1866, invokes the decreasing marginal utility of gain versus risk to explain why a rational person would not pay a large sum to play this game.

The utility y , which Bernoulli called the *emolumentum* (*gain*), is an important tool in economic analysis, since it provides a dynamic model of economic behavior: Buyers exchange money for goods or services of higher personal utility; sellers exchange goods and services for money of higher personal utility. If money, goods, and services did not have different utility for different people, no market could exist at all.¹¹ That idea is valid independently of the actual formula for utility given by Bernoulli, although, as far as measurements of psychological phenomena can be made, Bernoulli's assumption was extremely good. The physiologist Ernst Heinrich Weber (1795–1878) asked blindfolded subjects to hold weights, which he gradually increased, and to say when they noticed an increase in the weight. He found that the threshold for a noticeable difference was indeed inversely proportional to the weight. That is, if S is the perceived weight and W the actual weight, then $dS = k dW/W$, where dW is the smallest increment that can be noticed and dS the corresponding perceived increment. Thus he found exactly the law *assumed* by Bernoulli for perceived increases in wealth.¹² Utility is of vital importance to the insurance industry, which makes its profit by having a large enough stake to play “games” that resemble the Petersburg paradox.

One important concept was missing from the explanation of the Petersburg paradox. Granted that one should expect the “expected” value of a quantity depending on chance, how *confidently* should one expect it? The question of *dispersion* or *variance* of a random quantity lies beneath the surface here and needed to be brought out. (Variance is the square of the standard deviation.) It turns out that when the expected value is infinite, or even when the variance is infinite, no rational projections can be made. Since we live in a world of finite duration and finite resources, however, each game will be played only a finite number of times. It follows that every actual game has a finite expectation and variance and is subject to rational analysis using them.

36.8. LAPLACE

Although Pierre-Simon Laplace (1749–1827) is known primarily as an astronomer, he also developed a great deal of theoretical physics. (The partial differential equation satisfied by harmonic functions is named after him.) He also understood the importance of probabilistic methods in the analysis of data. In his *Théorie analytique des probabilités*, he proved that the distribution of the average of random observational errors that are uniformly distributed in an interval symmetric about zero tends to the normal distribution as the number of observations increases. Except for using the letter c where we now use e to denote the base of natural logarithms, he had what we now call the central limit theorem for independent uniformly distributed random variables.

¹⁰Pronounced “Keys.”

¹¹One feels the lack of this concept very strongly in the writing on economics by Aristotle and his followers, especially in their condemnation of the practice of lending money at interest, which ignores the utility of time.

¹²Weber's result was publicized by Gustave Theodor Fechner (1801–1887) and is now known as the Weber–Fechner law.

36.9. LEGENDRE

In a treatise on methods of determining the orbits of comets, published in 1805, Legendre dealt with the problem that frequently results when observation meets theory. Theory prescribes a certain number of equations of a particular form to be satisfied by the observed quantities. These equations involve certain parameters that are not observed, but are to be determined by fitting observations to the theoretical model. Observation provides a large number of empirical, approximate solutions to these equations, and thus normally provides a number of equations far in excess of the number of parameters to be chosen. If the law is supposed to be represented by a straight line, for example, only two constants are to be chosen. But the observed data will normally not lie on a line; instead, they may cluster around a line. How is the observer to choose canonical values for the parameters from the observed values of each of the quantities?

Legendre's solution to this problem is now a familiar technique. If the theoretical equation is $y = f(x)$, where $f(x)$ involves parameters α, β, \dots , and one has data points (x_k, y_k) , $k = 1, \dots, n$, sum the squares of the "errors" $f(x_k) - y_k$ to get an expression in the parameters

$$E(\alpha, \beta, \dots) = \sum_{k=1}^n (f(x_k) - y_k)^2,$$

and then choose the parameters so as to minimize E . For fitting with a straight line $y = ax + b$, for example, one needs to choose $E(a, b)$ given by

$$E(a, b) = \sum_{k=1}^n (ax_k + b - y_k)^2$$

so that

$$\frac{\partial E}{\partial a} = 0 = \frac{\partial E}{\partial b}.$$

36.10. GAUSS

Legendre was not the first to study the problem of determining the most likely value of a quantity x using the results of repeated measurements of it, say x_k , $k = 1, \dots, n$. In 1799 Laplace had tried the technique of taking the value x that minimizes the sum of the absolute errors¹³ $|x - x_k|$. But still earlier, in 1794 as shown by his diary and correspondence, the teenager Gauss had hit on the least-squares technique for the same purpose. As Reich (1977, p. 56) points out, Gauss did not consider this discovery very important and did not publish it until 1809. In 1816, Gauss published a paper on observational errors, in which he discussed the most probable value of a variable based on a number of observations of it. His discussion was much more modern in its notation than those that had gone before, and also much more

¹³This method has the disadvantage that one large error and many small errors count equally. The least-squares technique avoids that problem.

rigorous. He found the likelihood of an error of size x to be

$$\frac{h}{\sqrt{\pi}} e^{-h^2 x^2},$$

where h was what he called the *measure of precision*. He showed how to estimate this parameter by inverse-probability methods. In modern terms, $h = 1/(\sigma\sqrt{2})$, where σ is the standard deviation. This work brought the normal distribution into a standard form, and it is now often referred to as the *Gaussian distribution*.

36.11. PHILOSOPHICAL ISSUES

The notions of chance and necessity have often played a role in philosophical speculation; in fact, most books on logic are kept in the philosophy sections of libraries. Many of the mathematicians who have worked in this area have had a strong interest in philosophy and have speculated on what probability means. In so doing, they have come up against the same difficulties that confront natural philosophers when trying to explain how induction works. Like Pavlov's dogs and Skinner's pigeons (see Chapter 1), human beings tend to form expectations based on frequent, but not necessarily invariable conjunctions of events and seem to find it very difficult to suspend judgment and live with no belief where there is no evidence.¹⁴ Can philosophy offer us any assurance that proceeding by induction based on probability and statistics is any better than, say, divination? Are insurance companies acting on *pure faith* when they offer to bet us that we will survive long enough to pay them more money (counting the return on investment) in life insurance premiums than they will pay out when we die? If probability is a subjective matter, is subjectivity the same as arbitrariness?

What *is* probability, when applied to the physical world? Is it merely a matter of frequency of observation, and consequently objective? Or do human beings have some innate faculty for assigning probabilities? For example, when we toss a coin twice, there are four distinguishable outcomes: HH, HT, TH, TT. Are these four equally likely? If one does not know the order of the tosses, only three possibilities can be distinguished: two heads, two tails, and one of each. Should those be regarded as equally likely, or should we imagine that we do know the order and distinguish all four possibilities?¹⁵ Philosophers still argue over such matters. Siméon-Denis Poisson (1781–1840) seemed to be having it both ways in his *Recherches sur la probabilité des jugemens (Investigations into the Plausibility of Inferences)* when he wrote that

The probability of an event is the reason we have to believe that it has taken place, or that it will take place.

¹⁴In his *Formal Logic*, Augustus De Morgan imagined asking a person selected at random for an opinion whether the volcanoes—he meant craters—on the unseen side of the moon were larger than those on the side we can see. He concluded, “The odds are, that though he has never thought of the question, he has a pretty stiff opinion in three seconds.”

¹⁵If the answer to that question seems intuitively obvious, please note that in more exotic applications of statistics, such as in quantum mechanics, either possibility can occur. Fermions have wave functions that are antisymmetric, and they distinguish between HT and TH; bosons have symmetric wave functions and do not distinguish them.

and then immediately followed up with this:

The measure of the probability of an event is the ratio of the number of cases favorable to that event, to the total number of cases favorable or contrary.

In the first statement, he appeared to be defining probability as a subjective event, one's own *personal* reason, but then proceeded to make that reason an objective thing by assuming equal likelihood of all outcomes. Without some restriction on the universe of discourse, these definitions are not very useful. We do not know, for example, whether our automobile will start tomorrow morning or not, but if the probability of its doing so were really only 50% because there are precisely two possible outcomes, most of us would not bother to buy an automobile. Surely Poisson was assuming some kind of symmetry that would allow the imagination to assign equal likelihoods to the outcomes, and intending the theory to be applied only in those cases. Still, in the presence of ignorance of causes, equal probabilities seem to be a reasonable starting point. The law of entropy in thermodynamics, for example, can be deduced as a tendency for an isolated system to evolve to a state of maximum probability, and maximum probability means the maximum number of equally likely states for each particle.

36.12. LARGE NUMBERS AND LIMIT THEOREMS

The idea of the law of large numbers was stated imprecisely by Cardano and with more precision by James Bernoulli. To better carry out the computations involved in using it, De Moivre was led to approximate the binomial distribution with what we now realize was the normal distribution. He, Laplace, and Gauss all grasped with different degrees of clarity the principle (central limit theorem) that when independent measurements are averaged, their distribution density tends to resemble the bell-shaped curve.

The law of large numbers was given its name in the 1837 work of Poisson just mentioned. Poisson discovered an approximation to the probability of getting at most k successes in n trials, valid when n is large and the probability p is small. He thereby introduced what is now known as the *Poisson distribution*, in which the probability of k successes is given by

$$p_k = e^{-\lambda} \frac{\lambda^k}{k!}.$$

The Russian mathematician Chebyshev¹⁶ (1821–1894) introduced the concept of a random variable and its mathematical expectation. He is best known for his 1846 proof of the weak law of large numbers for repeated independent trials. That is, he showed that the probability that the actual proportion of successes will differ from the expected proportion by less than any specified $\varepsilon > 0$ tends to 1 as the number of trials increases. In 1867 he proved what is now called *Chebyshev's inequality*: *The probability that a random variable will assume a value more than [what is now called k standard deviations] from its mean is at most $1/k^2$.* This inequality was published by Chebyshev's friend and translator Irénée-Jules Bienaymé (1796–1878) and is sometimes called the *Chebyshev–Bienaymé inequality* (see

¹⁶Properly pronounced “Cheb-wee-SHAWF,” but more commonly “CHEB-ee-shev,” by English speakers.

Heyde and Seneta, 1977). This inequality implies the weak law of large numbers. In 1887, Chebyshev also stated the central limit theorem for independent random variables.

The extension of the law of large numbers to dependent trials was achieved by Chebyshev's student Andrei Andreevich Markov (1856–1922). The subject of dependent trials—known as *Markov chains*—remains an object of current research. In its simplest form it applies to a system in one of a number of states $\{S_1, \dots, S_n\}$ which at specified times may change from one state to another. If the probability of a transition from S_i to S_j is p_{ij} , the matrix

$$P = \begin{pmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{pmatrix}$$

is called the *transition matrix*. If the transition probabilities are the same at each stage, one can easily verify that the matrix power P^k gives the probabilities of the transitions in k steps.

PROBLEMS AND QUESTIONS

Mathematical Problems

36.1. We saw above that Cardano (probably) and Pascal and Leibniz (certainly) miscalculated some elementary probabilities. As an illustration of the counterintuitive nature of many simple probabilities, consider the following hypothetical games. (A casino could probably be persuaded to provide these games if there was enough public interest in them.) In Game 1 the dealer draws two randomly chosen cards from a deck on the table and looks at them. If neither card is an ace, the dealer shows them to the other players, and no game is played. The cards are replaced in the deck, the deck is shuffled, and the game begins again. If one card is an ace, players are not shown the cards, but are invited to bet against a fixed winning amount offered by the house that the other card is also an ace. What winning should the house offer (in order to break even in the long run) if players pay one dollar per bet?

In Game 2 the rules are the same, except that the game is played only when one of the two cards is the ace of hearts. What winning should the house offer in order to break even charging one dollar to bet? Why is this amount not the same as for Game 1?

36.2. Use the Maclaurin series for $e^{-(1/2)t^2}$ to verify that the series given by De Moivre, which was

$$\sqrt{\frac{2}{\pi}} \left(\frac{1}{0! \cdot 1 \cdot 2} - \frac{1}{1! \cdot 3 \cdot 4} + \frac{1}{2! \cdot 5 \cdot 8} - \frac{1}{3! \cdot 7 \cdot 16} + \cdots \right),$$

represents the integral

$$\frac{1}{\sqrt{2\pi}} \int_0^1 e^{-\frac{1}{2}t^2} dt,$$

which is the area under a standard normal (bell-shaped) curve above the mean, but by at most one standard deviation, as given in many tables.

- 36.3.** Radium-228 is an unstable isotope. Each atom of Ra-228 has a probability of 0.1145 (about 1 chance in 9, or about the probability of rolling a 5 with two dice) of decaying to form an atom of actinium within any given year. This means that the probability that the atom will survive the year as an atom of Ra-228 is $1 - 0.1145 = 0.8855$. Denote this “one-year survival” probability by p . Because any sample of reasonable size contains a huge number of atoms, that survival probability (0.8855) is the proportion of the weight of Ra-228 that we would expect to survive a year.

If you had one gram of Ra-228 to begin with, after one year you would expect to have $p = 0.8855$ grams. Each succeeding year, the weight of the Ra-228 left would be multiplied by p , so that after two years you would expect to have $p^2 = (0.8855)^2 = 0.7841$ grams. In general, after t years, if you started with W_0 grams, you would expect to have $W = W_0 p^t$ grams. Now push these considerations a little further and determine *how strongly* you can rely on this expectation. Recall Chebyshëv’s inequality, which says that the probability of being more than k standard deviations from the expected value is never larger than $(1/k)^2$. What we need to know to answer the question in this case is the standard deviation σ .

Assume that each atom decays at random, independently of what happens to any other atom. This independence allows us to think that observing our sample for a year amounts to a large number of “independent trials,” one for each atom. We test each atom to see if it survived as an Ra-228 atom or decayed into actinium. Let N_0 be the number of atoms that we started with. Assuming that we started with 1 gram of Ra-228, there will be $N_0 = 2.642 \cdot 10^{21}$ atoms of Ra-228 in the original sample.¹⁷ That is a very large number of atoms. The survival probability is $p = 0.8855$. For this kind of independent trial, as mentioned the standard deviation with N_0 trials is

$$\sqrt{N_0 p(1-p)} = \sqrt{\frac{p(1-p)}{N_0}} N_0.$$

We write the standard deviation in this odd-looking way so that we can express it as a fraction of the number N_0 that we started with. Since weights are proportional to the number of atoms, that same fraction will apply to the weights as well.

Put in the given values of p and N_0 to compute the fraction of the initial sample that constitutes one standard deviation. Since the original sample was assumed to be one gram, you can regard the answer as being expressed in grams. Then use Chebyshëv’s inequality to estimate the probability that the amount of the sample remaining will differ from the theoretically predicted amount by 1 millionth of a gram (1 microgram, that is, 10^{-6} grams)? [*Hint*: How many standard deviations is one millionth of a gram?]

Historical Questions

- 36.4.** From what real-life situations did the first mathematical analyses of probability arise?

¹⁷The number of atoms in one gram of Ra-228 is the *Avogadro number* $6.023 \cdot 10^{23}$ divided by 228.

- 36.5.** What new concepts were introduced in James Bernoulli's *Ars conjectandi*?
- 36.6.** What are the two best-known mathematical theorems about the probable outcome of a large number of random trials?

Questions for Reflection

- 36.7.** Consider the case of 200 men and 200 women applying to a university consisting of only two different departments, and assume that the acceptance rates are given by the following table.

	Men	Women
Department A	120/160	32/40
Department B	8/40	40/160

Observe that the admission rate for men in department A is $\frac{3}{4}$, while that for women is $\frac{4}{5}$. In department B the admission rate for men is $\frac{1}{5}$ and for women it is $\frac{1}{4}$. In both cases, the people making the admission decisions are admitting a higher proportion of women than of men. Yet the overall admission rate is 64% for men and only 36% for women. Explain this paradox in simple, nonmathematical language.

This paradox was first pointed out by George Udny Yule (1871–1951) in 1903. Yule produced a set of two 2×2 tables, each of which had no correlation, but produced a correlation when combined (see David and Edwards, 2001, p. 137). Yule's result was, for some reason, not given his name; but because it was publicized by Edward Hugh Simpson in 1951,¹⁸ it came to be known as *Simpson's paradox*.¹⁹ Simpson's paradox is a counterintuitive oddity, not a contradiction.

An example of it occurred in the admissions data from the graduate school of the University of California at Berkeley in 1973. These data raised some warning flags. Of the 12,763 applicants, 5232 were admitted, giving an admission rate of 41%. However, investigation revealed that 44% of the male applicants had been admitted and only 35% of the female applicants. There were 8442 male applicants, 3738 of whom were accepted, and 4321 female applicants, 1494 of whom were accepted. Simple chi-square testing showed that the hypothesis that these numbers represent a random deviation from a sex-independent acceptance rate of 41% was not plausible. There was unquestionably *bias*. The question was: Did this bias amount to discrimination? If so, who was doing the discriminating?

For more information on this case study and a very surprising conclusion, see "Sex bias in graduate admissions: Data from Berkeley," *Science*, **187**, 7 February 1975, 398–404. In that paper, the authors analyzed the very evident *bias* in admissions to look for evidence of *discrimination*. Since admission decisions are made by the individual departments, it seemed logical to determine which departments had a noticeably higher admission rate for men than for women. Surprisingly, the authors

¹⁸See "The interpretation of interaction in contingency tables," *Journal of the Royal Statistical Society*, Series B, **13**, 238–241.

¹⁹The name *Simpson's paradox* goes back at least to the article of C. R. Blyth, "On Simpson's paradox and the sure-thing principle," in the *Journal of the American Statistical Association*, **67** (1972), 364–366.

found only four such departments (out of 101), and the imbalance resulting from those four departments was more than offset by six other departments that had a higher admission rate for women.

- 36.8.** It is interesting that an exponential law of growth or decay can be associated with both completely deterministic models of growth (compound interest) and completely random models, as in the case of radioactive decay. Is it possible to assume physical laws that would make radioactive decay completely deterministic?
- 36.9.** In the Power Ball lottery that is played by millions of people in the United States every week, players are trying to guess which five of 59 numbered balls will drop out of a hopper, and which one of 39 others will drop out of a hopper. The jackpot is won by guessing all six correctly. Any combination of 5 numbers and 1 number is as likely to drop as any other. Hence, the number of possible combinations is

$$39 \binom{59}{5} = \frac{59 \cdot 58 \cdot 57 \cdot 56 \cdot 55 \cdot 39}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 195,249,054.$$

This is, needless to say, a rather large number, meaning that the odds are heavily against the player. To picture the odds more vividly, imagine that the winning combinations are the serial numbers on dollar bills, and that this \$195,249,054 consists of one-dollar bills laid end to end. They would cover more than 30,000 km, roughly the distance from the North Pole to the South Pole and then north again to the Equator. Even if the journey could all be made on level ground, it would require a vigorous walker 250 days to make it with no stops to rest. In that scenario, the gambler is going to bend down just once and hope to pick up the winning dollar bill.

Still, somebody does eventually win the Power Ball, every few weeks or months. Why is this fact not surprising? Is it a rational wager to spend a dollar to play the lottery if the prize goes above this stake, on the grounds that the expected gain is larger than the cost of playing?

Algebra from 1600 to 1850

By the mid-seventeenth century, the relation between the coefficients and roots of a general equation was understood, and it was conjectured that if you counted roots according to multiplicity and allowed complex roots, an equation of degree n would have n roots. Algebra had been consolidated to the point that the main unsolved problem, the solution of equations of degree higher than 4, could be stated simply and analyzed.

The solution of this problem took nearly two centuries, and it was not until the late eighteenth and early nineteenth centuries that enough insight was gained into the process of determining the roots of an equation from its coefficients to prove that arithmetic operations and root extractions were not sufficient for this purpose. Although the solution was a negative result, it led to the important concepts of modern algebra that we know as groups, rings, and fields; and these, especially groups, turned out to be applicable in many areas not directly connected with algebra. Also on the positive side, nonalgebraic methods of solving higher-degree equations were found, along with a criterion to determine whether or not a prescribed set of roots can be expressed algebraically in terms of the coefficients of the equation that they satisfy.

37.1. THEORY OF EQUATIONS

Viète understood something of the relation between the roots and the coefficients of some equations. His understanding was not complete, because he was not able to find all the roots. Before the connection could be made completely, there had to be a domain in which an equation of degree n would have n roots. Such a domain is called an *algebraically closed field*. Then the general connection between coefficients and roots could be made for quadratic, cubic, and quartic equations and generalized from there. The missing theorem was eventually to be called the *fundamental theorem of algebra*.¹

¹In his textbook on analytic function theory (1959, p. 24), Einar Hille (1894–1980) wrote that “modern algebraists are inclined to deny both its algebraic and its fundamental character.” In the context of its time, the theorem was both algebraic and fundamental. The fact that the complex numbers are algebraically closed depends on the topological properties of the complex plane, especially the fact that it is a connected set. That connectedness would not exist if not for nonalgebraic (transcendental) numbers such as π and e . That is the meaning of Hille’s statement that the algebraic closedness of the complex numbers is not an algebraic theorem.

37.1.1. Albert Girard

This fundamental theorem was first stated by Albert Girard (1595–1632), the editor of the works of Simon Stevin. In 1629 he wrote *L'invention nouvelle en l'algèbre* [*New Discovery (Invention) in Algebra*]. This work contained some of the unifying concepts that make modern algebra the compact, efficient system that it now is. One of these ideas is to regard the constant term as the coefficient of the zeroth power of the unknown. He introduced the notion of *factions* of a finite set of numbers. The first faction is the sum of the numbers, the second one is the sum of all products of two distinct numbers from the set, and so on. The last faction is the product of all the numbers, so that “there are as many factions as there are numbers given.” He noted that the number of terms in each faction could be found by using Pascal’s triangle.

Girard always regarded the leading coefficient as 1. Putting the equation into this form, he stated as a theorem (see, for example, Struik, 1986, p. 85) that “all equations of algebra receive as many solutions as the denomination [degree] of the highest form shows, except the incomplete, and the first faction of the solutions is equal to the number of the first mixed [that is, the coefficient of the power one less than the degree of the equation], their second faction is equal to the number of the second mixed, their third to the third mixed, and so on, so that the last faction is equal to the closure [product], and this according to the signs that can be observed in the alternate order.” This recognition that the coefficients of a polynomial are elementary symmetric polynomials in its zeros was the first ray of light at the dawn of modern algebra.

By “incomplete,” Girard seems to have meant equations with some terms missing. In some cases, he said, these may not have a full set of solutions. He gave the example of the equation $x^4 = 4x - 3$, whose solutions he gave as 1, 1, $-1 + \sqrt{-2}$, and $-1 - \sqrt{-2}$, showing that he realized the need to count both complex roots and multiple real roots for the sake of the general rule. It is not clear what connection he made between missing terms and a reduced number of solutions. If an equation $p(x) = 0$ with real coefficients has a pure imaginary solution $x = c\sqrt{-1}$, then $p(x)$ is divisible by $x^2 + c^2$, and there will be missing terms. But there may be terms missing even in an equation with a full set of solutions, for example, $x^4 - 13x^2 + 36$; and there may be no missing terms in an equation with no real solutions, such as $x^2 - x + 1 = 0$. He invoked the simplicity of the general rule as justification for introducing the multiple and complex roots, along with the fact that complex numbers provide solutions where otherwise none would exist.

37.1.2. Tschirnhaus Transformations

Every complex number has n th roots—exactly n of them except in the case of 0—that are also complex numbers. As a consequence, any formula for solving equations with complex coefficients that involves only the application of rational operations and root extractions starting with the coefficients will remain within the domain of complex numbers. This elementary fact led to the proposition stated by Girard, which we know as the fundamental theorem of algebra. Finding such a formula for equations of degree five and higher was to become a preoccupation of algebraists for the next two centuries.

By the year 1600, equations of degrees 2, 3, and 4 could all be solved, assuming that one could extract the cube root of a complex number (and that problem could not and cannot be reduced to purely algebraic operations on real numbers). The methods used to solve it—reducing the cubic to a quadratic equation in x^3 and reducing the quartic to the

resolvent cubic—suggest an inductive process in which the solution of an equation of degree n , say

$$x^n - a_1x^{n-1} + \cdots \mp a_{n-1}x \pm a_n = 0,$$

would be found by a substitution $y = x^{n-1} - b_1x^{n-2} + \cdots \pm b_{n-2}x \mp b_{n-1}$ with the coefficients b_1, \dots, b_{n-1} chosen so that the original equation becomes $y^n = C$. Here we have $n - 1$ coefficients b_k at our disposal and $n - 1$ coefficients a_1, \dots, a_{n-1} to be removed from the original equation. The program looks feasible. Something of the kind must have been the reasoning that led Ehrenfried Walther von Tschirnhaus (1652–1708) to the belief that he had discovered a general solution to all polynomial equations. In 1677 he wrote to Leibniz:

In Paris I received some letters from Mr. Oldenburg, but from lack of time have not yet been able to write back that I have found a new way of determining the irrational roots of all equations. . . The entire problem reduces to the following: We must be able to remove all the middle terms from any equation. When that is done, and as a result only a single power and a single known quantity remain, one need only extract the root.

Tschirnhaus claimed that the the middle terms (the a_k above) would be eliminated by a polynomial of the sort just discussed, provided that the b_k are suitably chosen. Such a change of variable is now called a *Tschirnhaus transformation*. If a Tschirnhaus transformation could be found for the *general* equation of degree n , and a formula existed for solving the *general* equation of degree $n - 1$, the two could be combined to generate a formula for solving the general equation of degree n . At the time, there was not even a Tschirnhaus transformation for the cubic equation. Tschirnhaus was to provide one.

He illustrated his transformation using the example $x^3 - qx - r = 0$. Taking $y = x^2 - ax - b$, he noted that y satisfied the equation²

$$y^3 + (3b - 2q)y^2 + (3b^2 + 3ar - 4qb + q^2 - a^2q)y + (b^3 - 2qb^2 + 3bar + q^2b - aqr - a^2qb + a^3r - r^2) = 0.$$

He eliminated the square term by choosing $b = 2q/3$, then removed the linear term by solving for a in the quadratic equation

$$qa^2 - 3ra + q^2/3 = 0.$$

In this way, he had found at the very least a second solution of the general cubic equation, independent of the solution given by Cardano. And, what is more important, he had indicated a plausible way by which any equation might be solved. If it worked, it would prove that every polynomial equation could be solved using rational operations and root extractions, thereby proving at the same time that the complex numbers are algebraically closed. Unfortunately, detailed examination of the problem revealed difficulties that Tschirnhaus had apparently not noticed at the time of his letter to Leibniz.

²This equation can be derived by Seki Kōwa's method of *tatami* (folding), which makes it possible to express x as a fractional-linear function of y , and hence y as a fractional-linear function of x .

When folding is used twice with the two polynomial equations $p_n(x) = 0$ and $y = p_{n-1}(x)$, where p_n is of degree n and p_{n-1} of degree $n - 1$, the polynomial that remains, just as in the case $n = 3$, contains a constant term and a linear term in x whose coefficients are linear functions of y . Those two terms would make it possible to express x as a fractional-linear function of y . Unfortunately, this polynomial also contains terms of degrees up to $n - 2$. Those terms can be removed by a suitable choice of parameters in $p_{n-1}(x)$, but doing so requires fixing all but 2 of the coefficients. As a result, it is not in general possible to remove more than two of the coefficients in the resulting equation of degree n in this way. Only in the case of a cubic does that elimination produce a pure equation. The process may, however, work for a *particular* equation of higher degree. Leibniz was not convinced. He wrote to Tschirnhaus,

I do not believe that [your method] will be successful for equations of higher degree, except in special cases. I believe that I have a proof for this. [Kracht and Kreyszig, 1990, p. 27]

Tschirnhaus' method had intuitive plausibility: If there existed an algorithm for solving all equations, that algorithm should be a procedure like the Tschirnhaus transformation. Because the method does *not* work, the thought suggests itself that there may be equations that cannot be solved algebraically. The work of Tschirnhaus and Girard had produced two important insights into the general problem of polynomial equations: (1) The coefficients are symmetric functions of the roots; (2) solving the equation should be a matter of finding a sequence of operations that would eliminate coefficients until a pure equation $y^n = C$ was obtained. Since the problem was still unresolved, still more new insights were needed.

To explain the most important of these new insights, let us consider what Girard's result means when applied to Cardano's solution of the cubic $y^3 + py = q$. If the roots of this equation are r , s , and t , then $p = st + tr + rs$, $q = rst$, $t = -r - s$, since the coefficient of y^2 is zero. The sequence of operations implied by Cardano's formula is

$$\begin{aligned} u &= \frac{p}{3}; & v &= \frac{q}{2}; \\ a &= \sqrt{u^3 + v^2}; \\ y &= \sqrt[3]{v + a} + \sqrt[3]{v - a}. \end{aligned}$$

Girard's work implies that the quantity a , which is an *irrational* function of the coefficients p and q , is a *rational* function of the roots r , s , and t :

$$a = \pm \frac{i}{\sqrt{108}}(r - s)(s - t)(t - r);$$

that is, it does not involve taking the square root of any expression containing a root.

37.1.3. Newton, Leibniz, and the Bernoullis

In the 1670s, Newton wrote a textbook of algebra called *Arithmetica universalis*, which was published in 1707, in which he stated more clearly and generally than Girard had done the relation between the coefficients and roots of a polynomial. Moreover, he showed that symmetric polynomials in the roots could be expressed as polynomials in the coefficients by giving a set of rules that are still known by his name.

Another impetus toward the fundamental theorem of algebra came from calculus. The well-known method known as partial fractions for integrating a quotient of two polynomials reduces all such problems to the purely algebraic problem of factoring the denominator. It is not immediately obvious that the denominator can be factored into linear and quadratic real factors; that is the content of the fundamental theorem of algebra. John Bernoulli asserted in a paper in the *Acta eruditorum* in 1702 that such a factoring was always possible, and therefore all rational functions could be integrated. Leibniz did not agree, arguing that the polynomial $x^4 + a^2$, for example, could not be factored into quadratic factors over the reals. Here we see a great mathematician being misled by following a method. He recognized that the factorization had to be $(x^2 + a^2\sqrt{-1})(x^2 - a^2\sqrt{-1})$ and that the first factor should therefore be factored as $(x + a\sqrt{-\sqrt{-1}})(x - a\sqrt{-\sqrt{-1}})$ and the second factor as $(x + a\sqrt{\sqrt{-1}})(x - a\sqrt{\sqrt{-1}})$, but he did not realize that these factors could be combined to yield $x^4 + a^2 = (x^2 - \sqrt{2}ax + a^2)(x^2 + \sqrt{2}ax + a^2)$. It was pointed out by Nicholas Bernoulli I (1687–1759) in the *Acta eruditorum* of 1719 (three years after the death of Leibniz) that this last factorization was a consequence of the identity $x^4 + a^4 = (x^2 + a^2)^2 - 2a^2x^2$.

37.2. EULER, D'ALEMBERT, AND LAGRANGE

The eighteenth century saw considerable progress in the understanding of equations in general and the procedures needed to solve them. Much of this new understanding came from the two men who dominated mathematical life in that century, Euler and Lagrange.

37.2.1. Euler

In his 1749 paper “Recherches sur les racines imaginaires des équations” (“Investigations into the imaginary roots of equations”), devoted to equations whose degree is a power of 2 and published in the memoirs of the Berlin Academy, Euler showed that when the coefficients of a polynomial are real, its roots occur in conjugate pairs, and therefore produce irreducible real quadratic factors of the form $(x - a)^2 + b^2$. In this paper, Euler argued that every polynomial of degree 2^n with real coefficients can be factored as a product of two polynomials of degree 2^{n-1} with real coefficients. In the course of the proof, Euler presented the germ of an idea that was to have profound consequences. In showing that a polynomial of degree 8 could be written as a product of two polynomials of degree 4, he assumed that the coefficient of x^7 was made equal to zero by means of a linear substitution. The remaining polynomial $x^8 - ax^6 + bx^5 - cx^4 - dx^2 + ex - f$ was then to be written as a product

$$(x^4 - ux^3 + \alpha x^2 + \beta x + \gamma)(x^4 + ux^3 + \delta x^2 + \varepsilon x + \zeta).$$

Euler noted that since u was the sum of four roots of the equation, it could assume (potentially) 70 values (the number of combinations of eight things taken four at a time), and its square would satisfy an equation of degree 35.

In the 1749 paper, Euler also conjectured that the roots of an equation of degree higher than 4 cannot be constructed by applying a finite number of algebraic operations to the coefficients. This was the first explicit statement of such a conjecture.

In his 1762 paper “De resolutione aequationum cuiusque gradus” (“On the solution of equations of any degree”), published in the *Commentarii* of the Petersburg Academy, Euler tried a different approach,³ assuming a solution of the form

$$x = w + A\sqrt[n]{v} + B\sqrt[n]{v^2} + \cdots + Q\sqrt[n]{v^{n-1}},$$

where w is a real number and v and the coefficients A, \dots, Q are to be found by a procedure resembling a Tschirnhaus transformation. This approach was useful for equations of degree 2^n , but fell short of being a general solution of all polynomial equations.

37.2.2. D’Alembert

Euler’s contemporary and correspondent d’Alembert tried to prove that all polynomials could be factored into linear and quadratic factors in order to prove that all rational functions could be integrated by partial fractions. In the course of his argument he assumed that any algebraic function could be expanded in a series of fractional powers of the independent variable. While Euler was convinced by this proof, he also wrote to d’Alembert to say that this assumption would be questioned (Bottazzini, 1986, pp. 15–18).

37.2.3. Lagrange

In 1770, Lagrange made a survey of the methods known up to his time for solving general equations. He devoted a great deal of space to a preliminary analysis of the cubic and quartic equations. In particular, he was intrigued by the fact that the resolvent equation, which he called the *reduced* equation (*équation réduite*), for the cubic was actually an equation of degree 6 that just happened to be quadratic in the third power of the unknown. He showed that if the roots of the cubic equation $x^3 + px = q$ being solved were a, b , and c , then a root of the resolvent would be

$$y = \frac{a + \alpha b + \alpha^2 c}{3},$$

where $\alpha^3 = 1$, $\alpha \neq 1$. He argued that since the original equation was symmetric in a, b , and c , the resolvent would have to admit this y as a root, no matter how the letters a, b , and c were permuted. It therefore followed that the resolvent would in general have six different roots. (Note, however, that y^3 assumes only two values under these permutations, and therefore satisfies a quadratic equation whose coefficients are rational functions of p and q that can be computed by an algorithm.)

For the quartic equation with roots a, b, c , and d , he showed that the resolvent cubic equation would have a root

$$t = \frac{ab + cd}{2}.$$

³This approach was discovered independently by Etienne Bézout (1730–1783).

Since this expression could assume only three different values when the roots were permuted—namely, half of $ab + cd$, $ac + bd$, or $ad + bc$ —it would have to satisfy an equation of degree three with coefficients expressible in terms of those of the original equation.

Proceeding to equations of fifth degree, Lagrange examined the only methods proposed up to that time, by Tschirnhaus and Euler–Bézout, and showed that the resolvent to be expected in all cases would be of degree 24. Pointing out that even Tschirnhaus, Euler, and Bézout themselves had not seriously attacked equations of degree five or higher, nor had anyone else tried to extend their methods, he said, “It is therefore greatly to be desired that one could estimate *a priori* the success that is to be expected in applying these methods to degrees higher than the fourth.” He then set out to provide proof that, in general, one could not expect the resolvent equation to reduce to lower degree than the original equation in such cases, at least using the methods mentioned.

To prove his point, Lagrange analyzed the method of Tschirnhaus from a more general point of view. For cubic and quartic equations, in which only two coefficients needed to be eliminated (the linear and quadratic terms in the cubic, the linear and cubic terms in the quartic), the substitution $y = x^2 + ax + b$ would always work, since the elimination procedure resulted in linear and quadratic expressions in a and b in the coefficients that needed to be eliminated. Still, as Lagrange remarked, that meant two pairs of possible values (a, b) and hence really two cubic resolvents to be solved. The resolvent was therefore once again an equation of degree 6, which happened to be a quadratic polynomial in the cube of the variable. He noted what must be an ominous sign for those hoping to solve all algebraic equations by algebraic methods: The construction of the coefficients in the resolvent for an equation of degree n appeared to require solving $n - 1$ equations in $n - 1$ unknowns, of degrees 1, 2, . . . , $n - 1$, so that eliminating the variable x in these equations therefore led to an expression for x that was of degree $(n - 1)!$ in y , and hence to a resolvent equation of degree $n!$ in y .

Lagrange summed up his analysis as follows:

To apply, for example, the method of Tschirnhaus to the equation of degree 5, one must solve four equations in four unknowns, the first being of degree 1, the second of degree 2, and so on. Thus the final equation resulting from the elimination of three of these unknowns will in general be of degree 24. But apart from the immense amount of labor needed to derive this equation, it is clear that after finding it, one will be hardly better off than before, unless one can reduce it to an equation of degree less than 5; and if such a reduction is possible, it can only be by dint of further labor, even more extensive than before.

The technique of counting the number of different values the root of the resolvent will have when the roots of the original equation are permuted among themselves was an important clue in solving the problem of the quintic.

37.3. THE FUNDAMENTAL THEOREM OF ALGEBRA AND SOLUTION BY RADICALS

The question of the theoretical existence of roots was settled on an intuitive level in the 1799 dissertation of Gauss. Gauss distinguished between the abstract *existence* of a root, which he proved, and an algebraic *algorithm* for finding it, the existence of which he doubted. He

pointed out that attempts to prove the existence of a root and any possible algorithm for finding it must assume the possibility of extracting the n th root of a complex number. He also noted the opinion, first stated by Euler, that no algebraic algorithm existed for solving the general quintic.

The reason we say that the existence of roots was settled only on the intuitive level is that the proof of the fundamental theorem of algebra is as much topological as algebraic. The existence of real roots of an equation of odd degree with real coefficients seems obvious since a real polynomial of odd degree tends to oppositely signed infinities as the independent variable ranges from one infinity to the other. It thus follows by connectivity that it must assume a zero at some point. Gauss' proof of the existence of complex roots was similar. Much of what he was doing was new at the time, and he had to explain it in considerable detail. For that reason, he preferred to use only real-variable methods, so as not to raise any additional doubts with the use of complex numbers. In fact, he stated his purpose in that way: to prove that every polynomial with real coefficients has a complete factorization into linear and quadratic real polynomials. (It was noted above that d'Alembert had proposed a proof of this theorem, but Gauss found it defective, since it conflated an infimum with a minimum.)

The complex-variable background of the proof is obvious nowadays, and Gauss admitted that his lemmas were normally proved using complex numbers. The steps were as follows. First, considering the equation $z^m + Az^{m-1} + Bz^{m-2} + \cdots + Kz^2 + Lz + M = 0$, where all coefficients A, \dots, M were real numbers,⁴ taking $z = r(\cos \varphi + i \sin \varphi)$ and using the relation $z^m = r^m(\cos m\varphi + i \sin m\varphi)$, one can see that finding a root amounts to setting the real and imaginary parts equal to zero simultaneously, that is, finding r and φ such that

$$\begin{aligned} r^m \cos m\varphi + Ar^{m-1} \cos(m-1)\varphi + \cdots + Kr^2 \cos 2\varphi + Lr \cos \varphi + M &= 0, \\ r^m \sin m\varphi + Ar^{m-1} \sin(m-1)\varphi + \cdots + Kr^2 \sin 2\varphi + Lr \sin \varphi &= 0. \end{aligned}$$

What remained was to show that there actually were points where the two curves intersected. For that purpose, Gauss divided both equations by r^m and argued that for large values of r the two functions must have zeros near the zeros of $\cos m\varphi = 0$ and $\sin m\varphi = 0$, respectively. That would mean that on a sufficiently large circle, each would have $2m$ zeros; and moreover, the zeros of the first curve, being near the points with polar angles $(k + 1/2)\pi/m$, must separate those of the second, which are near the points with polar angles $k\pi/m$. Then, arguing that the portion of each curve inside the disk of radius r was connected, he said that it was obvious that one could not join all the pairs from one set and all the pairs from the other set using two curves that do not intersect.

Gauss was uneasy about the intuitive aspect of the proof. During his lifetime he gave several other proofs of the theorem that he regarded as more rigorous.

37.3.1. Ruffini

As it turned out, Gauss had no need to publish his own research on the quintic equation. In the very year in which he wrote his dissertation, the first claim of a proof that it is impossible to find a formula for solving all quintic equations by algebraic operations was made by the Italian physician Paolo Ruffini (1765–1822). Ruffini's proof was based on Lagrange's

⁴This restriction involves no loss of generality (see Problem 37.1 below).

count of the number of values a function can assume when its variables are permuted.⁵ The principles of such a proof were gradually coming into focus. Newton’s principle that every symmetric polynomial in the roots of a polynomial can be expressed as a function of its coefficients, proved by Edward Waring (1736–1798), was an important step, as was the idea of counting the number of different values a rational function of the roots can assume. To get the general proof, it was necessary to show that the root extractions performed in the course of a hypothetical solution would also be rational functions of the roots. That this is the case for quadratic and cubic equations is not difficult to see, since the quadratic formula for solving $x^2 - (r_1 + r_2)x + r_1r_2 = 0$ involves taking only one square root:

$$\sqrt{(r_1 + r_2)^2 - 4r_1r_2} = \sqrt{(r_1 - r_2)^2} = \pm(r_1 - r_2).$$

Similarly, the Cardano formula for solving $y^3 + (r_1r_2 + r_2r_3 + r_3r_1)y = r_1r_2r_3$, where $r_1 + r_2 + r_3 = 0$, involves taking

$$\sqrt{\frac{(r_1r_2 + r_2r_3 + r_3r_1)^3}{27} + \frac{(r_1r_2r_3)^2}{4}} = \sqrt{\frac{-1}{108} \left((r_1 - r_2)(2r_1^2 + 5r_1r_2 + 2r_2^2) \right)},$$

followed by extraction of the cube roots of the two numbers

$$\frac{i}{3\sqrt{3}}(r_1 + \omega r_2)^3 \quad \text{and} \quad \frac{i}{3\sqrt{3}}(r_1 + \omega^2 r_2)^3,$$

where $\omega = -1/2 + i\sqrt{3}/2$ is a complex cube root of 1. All of these radicals are consequently rational (but not symmetric) functions of the roots.

37.3.2. Cauchy

Although Ruffini’s proof was not fully accepted by his contemporaries, it was endorsed many years later by Cauchy. In 1812, Cauchy wrote a paper entitled “Essai sur les fonctions symétriques” in which he proved the crucial fact that a polynomial in 5 variables that assumes fewer than 5 values when its variables are permuted assumes at most two values. In 1815 he published this result.

Cauchy gave credit to Lagrange, Alexandre Théophile Vandermonde (1735–1796), and Ruffini for earlier work in this area. Vandermonde, in particular, exhibited the Vandermonde determinant

$$\det \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \dots & \vdots & \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix} = -(x_1 - x_2)(x_1 - x_3) \cdots (x_1 - x_n)(x_2 - x_3) \cdots (x_2 - x_n) \cdots (x_{n-1} - x_n),$$

⁵An exposition of Ruffini’s proof, clothed in modern terminology that Ruffini would not have recognized, can be found in the paper of Ayoub (1980).

which assumes only two values, since interchanging two variables transposes the corresponding rows of the determinant and hence reverses the sign of the determinant.

37.3.3. Abel

Cauchy's work had a profound influence on two young geniuses whose lives were destined to be very short. The first of these, Abel, believed in 1821 that he had succeeded in solving the quintic equation. He sent his solution to the Danish mathematician Ferdinand Degen (1766–1825), who asked him to provide a worked-out example of a quintic equation that could be solved by Abel's method. While working through the details of an example, Abel realized his mistake. In 1824, he constructed an argument to show that such a solution was impossible and had the proof published privately. A formal version was published in the *Journal für die reine und angewandte Mathematik* in 1826. Abel was aware of Ruffini's work, and mentioned it in his argument. He attempted to fill in the gap in Ruffini's work with a proof that the intermediate radicals in any supposed solution by formula can be expressed as rational functions of the roots.

Abel's idea was that if some finite sequence of rational operations and root extractions applied to the coefficients produces a root of the equation

$$x^5 - ax^4 + bx^3 - cx^2 + dx - e = 0,$$

the final result must be expressible in the form

$$x = p + R^{\frac{1}{m}} + p_2 R^{\frac{2}{m}} + \cdots + p_{m-1} R^{\frac{m-1}{m}},$$

where p, p_2, \dots, p_{m-1} , and R are also formed by rational operations and root extractions applied to the coefficients, m is a prime number,⁶ and $R^{1/m}$ is not expressible as a rational function of the coefficients $a, b, c, d, e, p, p_2, \dots, p_{m-1}$.⁷ By straightforward reasoning on a system of linear equations for the coefficients p_j , he was able to show that R is a symmetric function of the roots, and hence that $R^{1/m}$ can assume the same m different values, no matter how the roots are permuted. Moreover, since there are $5!$ permutations of the roots and m is a prime, it followed that $m = 2$ or $m = 5$, the case $m = 3$ having been ruled out by Cauchy. The hypothesis that $m = 5$ led to an equation in which the left-hand side assumed only five values while the right-hand side assumed 120 values as the roots were permuted. Then the hypothesis $m = 2$ led to a similar equation in which one side assumed 120 values and the other only 10. Abel concluded that the hypothesis that there exists an algorithm for solving the equation was incorrect.

The standard version of the history of mathematics credits Abel with being "the" person who proved the impossibility of solving the quintic equation. But according to Ayoub (1980, p. 274), in 1832 the Prague Scientific Society declared the proofs of Ruffini and Abel unsatisfactory and offered a prize for a correct proof. The question was investigated

⁶Extracting any root is tantamount to the sequential extraction of prime roots. Hence every root extraction in the hypothetical process of solving the equation can be assumed to be the extraction of a prime root.

⁷Abel incorporated the apparently missing coefficient p_1 into R here, since he saw no loss of generality in doing so. A decade later, William Rowan Hamilton pointed out that doing so might increase the index of the root that needed to be extracted, since p_1 might itself require the extraction of an m th root.

by Hamilton in a report to the Royal Society in 1836 and published in the *Transactions of the Royal Irish Academy* in 1839. Hamilton's report was so heavily laden with subscripts and superscripts bearing primes that only the most dedicated reader would attempt to understand it, although Felix Klein (1884) was later to describe it as being "as lucid as it is voluminous." The proof was described by the American number theorist and historian of mathematics Leonard Eugene Dickson as "a very complicated reconstruction of Abel's proof." Hamilton regarded the problem of the solvability of the quintic as still open. He wrote:

[T]he opinions of mathematicians appear to be not yet entirely agreed respecting the possibility or impossibility of expressing a root as a function of the coefficients by any finite combination of radicals and rational functions.

The verdict of history has been that Abel's proof, suitably worded, is correct. Ruffini also had a sound method (see Ayoub, 1980), but needed to make certain subtle distinctions that were noticed only after the problem was better understood. By the end of the nineteenth century, Klein (1884) referred to "the proofs of *Ruffini* and *Abel*, by which it is established that a solution of the general equation of the fifth degree by extracting a finite number of roots is impossible."

Besides his impossibility proof, Abel made positive contributions to the solution of equations. He generalized the work of Gauss on the cyclotomic (circle-splitting) equation $x^n + x^{n-1} + \cdots + x + 1 = 0$, which had led Gauss to the construction of the regular 17-sided polygon. Abel showed that if every root of an equation could be generated by applying a given rational function successively to a single (primitive) root, the equation could be solved by radicals. Any two permutations that leave this function invariant necessarily commute with each other. As a result, nowadays any group whose elements commute is called an *abelian* group.

37.3.4. Galois

More light was shed on the solution of equations by the work of Abel's contemporary Evariste Galois (1811–1832), a volatile young man who did not live to become even mature. As is well known, he died at the age of 20 in a duel fought with one of his fellow republicans.⁸

The concepts of group, ring, and field that make modern algebra the beautiful subject that it is grew out of the work of Abel and Galois, but neither of these two short-lived geniuses had a full picture of any of them. Where we now talk easily about *algebraic and transcendental field extensions* and regard the general equation of degree n over a field F as $x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n = 0$, where each a_j is transcendental over F , Galois had to explain that the concept of a rational function was relative to what was given. For an equation with numerical coefficients, a rational function was simply a quotient of two

⁸The word *republican* (*républicain*) is being used in its French sense, of course, not the American sense. It is the opposite of *royaliste* or *légitimiste*. (Cauchy was a *légitimiste*, who went into exile with the King after the 1848 revolution.) There are murky details about the duel, but it appears that the gun Galois used was not loaded, probably because he did not wish to kill a comrade-in-arms. It may be that the combatants had agreed to let fate determine the outcome and each picked up a weapon, not knowing which of the two guns was loaded. The cause of the duel is also not entirely clear. The notes that Galois left behind seem to imply that he felt it necessary to warn his friends about what he considered to be the wiles of a certain young woman by whom he felt betrayed, and they felt obliged to defend her honor against his remarks.

polynomials with numerical coefficients, while if the equation had letters as coefficients, a rational function meant a quotient of two polynomials whose coefficients were rational functions of the coefficients of the equation. Even the concept of a group, which is associated with Galois, is not stated formally in any of his work. He does use the word *group* frequently in referring to a set of permutations of the roots of an equation, and he uses the properties that we associate with a group: the composition of permutations. However, it is clear from his language that what makes a set of permutations a group is that *all of them have the same effect on certain rational functions of the roots*. In particular, when what we now call a group is decomposed into cosets over a subgroup, Galois refers to the cosets as groups, since any two elements of a given coset have the same effect on the rational functions. He says that a group, in this sense, may begin with any permutation at all, since there is no need to specify any natural order of the roots.

Besides the shortness of their lives, Abel and Galois had another thing in common: neglect of their achievements by the Paris Academy of Sciences. Abel's most brilliant paper on general algebraic integrals (now called *abelian integrals* at the suggestion of Abel's rival Jacobi) was lost for over a decade until Jacobi, in 1841, insisted on its publication. As for Galois, he had been expelled from the École Normale because of his republican activities and had been in prison. He left a second paper among his effects, which was finally published in 1846. It had been written in January 1831, 17 months before his death, and it contained the following plaintive preface:

The attached paper is excerpted from a work that I had the honor to present to the Academy a year ago. Since this work was not understood, and doubt was cast on the propositions that it contains, I have had to settle for giving the general principles and *only one* application of my theorie in systematic order. I beg the referees at least to read these few pages with attention. [Picard, 1897, p. 33]

In a letter to a friend written the night before the duel in which he died, Galois showed that he had gone still further into this subject, making the distinction between proper and improper decompositions of the group of an equation, that is, the distinction we now make between normal and nonnormal subgroups.

The ideas of Abel and Galois were developed further by Laurent Wantzel (1814–1848) and Enrico Betti (1823–1892). Using reasoning about the roots of equations similar to that of Abel, Wantzel was able to prove (1837) that it is impossible to double the cube or trisect the angle using ruler and compass. More generally, he showed that any complex number that can be located using a straightedge and compass starting from the numbers 0 and 1 must satisfy an equation obtained by substituting one quadratic polynomial inside another a finite number n of times in such a way that the final polynomial of degree 2^n is irreducible over the rational numbers. At the time, no one knew whether π could be the root of such a polynomial, and so the question of squaring the circle remained open for several more decades. Later, Wantzel gave a thorough classification of the roots of equations with rational coefficients, and in the process showed (1843, p. 125) that there is no algebraic algorithm that can be applied to the coefficients of a cubic equation with three real roots and yield a root without involving complex numbers at some intermediate stage. In 1845, he “cleaned up” Abel's proof that it is impossible to solve equations of degree five or higher in radicals (as Hamilton had already done nine years earlier). Nowadays, all of Wantzel's results are proved using Galois theory.

In 1852, Betti published a series of theorems elucidating the theory of solvability by radicals. In this way, group theory proved to be the key not only to the solvability of equations but to the full understanding of classical problems. When Ferdinand Lindemann (1852–1939) proved in 1881 that π is a transcendental number, it followed that no ruler-and-compass quadrature of the circle was possible.

The proof that the general quintic equation of degree 5 was not solvable by radicals naturally raised two questions: (1) How *can* the general quintic equation of degree 5 be solved? (2) Which particular quintic equations *can* be solved by radicals? These questions required some time to answer. Eventually, Charles Hermite⁹ (1822–1902) showed how to use elliptic integrals to solve the general quintic equation. The algebraic algorithm for solving a solvable quintic equation depends on the Galois group of the equation. Using modern computer algebra methods, mathematicians have produced a complete solution of this problem.¹⁰ An early summary of results in this direction was the famous book by Felix Klein on the icosahedron (1884). A study of the theory of solvability of equations of all degrees, with historical reference, can be found in the book of R. Bruce King (1996).

PROBLEMS AND QUESTIONS

Mathematical Problems

- 37.1.** Prove that if every polynomial with real coefficients has a zero in the complex numbers, then the same is true of every polynomial with complex coefficients. To get started, let $p(z) = z^n + a_1z^{n-1} + \cdots + a_{n-1}z + a_n$ be a polynomial with complex coefficients a_1, \dots, a_n . Consider the polynomial $q(z)$ of degree $2n$ given by $q(z) = p(z)\overline{p(\bar{z})}$, where the overline indicates complex conjugation. This polynomial has real coefficients, and so by hypothesis has a complex zero z_0 .
- 37.2.** You are familiar with the fact that for any two polynomials $p(x)$ and $q(x)$, there exist other polynomials $a(x)$ and $r(x)$ such that $p(x) = a(x)q(x) + r(x)$, where $r(x)$ is of lower degree than $q(x)$. (This is the familiar long-division algorithm for polynomials.) The operation can be repeated using $q(x)$ and $r(x)$ in place of $p(x)$ and $q(x)$, eventually producing a remainder of maximal degree that divides both $p(x)$ and $q(x)$. This is the Euclidean algorithm discussed in Section 9.1 of Chapter 9. Because of this algorithm, it follows that every polynomial $p(x)$ can be factored in only one way (up to constant multiples) as a product of irreducible polynomials, that is, polynomials that cannot be divided by any non-constant polynomial of lower degree. This is because an irreducible polynomial is *prime* in the sense that if it divides a product of two polynomials, it must divide one of the factors. Use this fact to prove that since $p(x)$ is divisible by $x - a$ if $p(a) = 0$, it follows that $p(x)$ cannot have a number of roots that exceeds its degree.
- 37.3.** Show that complex numbers of the form $m + n\sqrt{2}i$, where m and n are integers can be added, subtracted, multiplied, and divided with a remainder, in the sense that given

⁹Hermite was also the first person to prove that e is a transcendental number, in 1858.

¹⁰See the paper by D. S. Dummit “Solving solvable quintics,” in *Mathematics of Computation*, **57** (1991), No. 195, 387–401.

$m + n\sqrt{2}i$ and $p + q\sqrt{2}i$, we can find $a + b\sqrt{2}i$ and $r + s\sqrt{2}i$ such that

$$m + n\sqrt{2}i = (p + q\sqrt{2}i)(a + b\sqrt{2}i) + (r + s\sqrt{2}i),$$

where $r + s\sqrt{2}i$ is smaller than $p + q\sqrt{2}i$ in the sense that the norm $N(r + s\sqrt{2}i)$, defined as $r^2 + 2s^2$, satisfies $N(r + s\sqrt{2}i) < N(p + q\sqrt{2}i)$. It follows from this result that an irreducible element of this form is prime in the sense of Problem 37.2, and hence that there is only one factorization (except for signs) for any number of this form. Does this result continue to hold if $\sqrt{2}$ is replaced by $\sqrt{3}$? [Hint: The complex quotient

$$c + d\sqrt{2}i = \frac{m + n\sqrt{2}i}{p + q\sqrt{2}i} = \frac{mp + 2nq}{p^2 + 2q^2} + \frac{(np - mq)\sqrt{2}i}{p^2 + 2q^2}$$

is not of the required form because c and d are not necessarily integers. However, there are integers a and b such that $|a - c| \leq \frac{1}{2}$ and $|b - d| \leq \frac{1}{2}$. Write $c = a + \delta$, $d = b + \varepsilon$, where $|\delta| \leq \frac{1}{2}$ and $|\varepsilon| \leq \frac{1}{2}$, and show that the difference $m + n\sqrt{2}i - (p + q\sqrt{2}i)(a + b\sqrt{2}i) = r + s\sqrt{2}i$ satisfies the required inequality. For the second question, note that $4 = (1 + \sqrt{3}i)(1 - \sqrt{3}i) = 2 \cdot 2$.]

Historical Questions

- 37.4.** What advance in clarity concerning polynomial equations is due to Girard?
37.5. What was the conclusion of Lagrange's detailed analysis of the general algebraic equation?
37.6. How did the results of Abel and Galois bring a measure of completeness to the search for an algebraic formula to solve each equation?

Questions for Reflection

- 37.7.** It is possible to define a multiplication on four-dimensional space by regarding the first coordinate as a real number and the last three as a vector. In other words, we can write formally

$$A = (a, a_1, a_2, a_3) = a + \alpha,$$

where a on the right stands for $(a, 0, 0, 0)$ and α for $(0, a_1, a_2, a_3)$. Addition of these *quaternions*, as they are called, is simple: $A + B = (a + b) + (\alpha + \beta)$, where of course $B = b + \beta = (b, b_1, b_2, b_3)$. Multiplication is a little trickier, and we define $A \times B = (ab - \alpha \cdot \beta) + (a\beta + b\alpha + \alpha \times \beta)$, where $\alpha \cdot \beta$ corresponds to the dot product $(a_1b_1 + a_2b_2 + a_3b_3, 0, 0, 0)$, and $\alpha \times \beta$ is the cross product $(0, a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1)$. It is not difficult to verify that $1 = (1, 0, 0, 0)$ has the property $1 \times A = A \times 1 = A$ for all quaternions A and that the quaternion conjugate $\bar{A} = a - \alpha$ satisfies $A \times \bar{A} = \bar{A} \times A = |A|^2 = (a^2 + a_1^2 + a_2^2 + a_3^2, 0, 0, 0)$, which is identified with the real number $a^2 + a_1^2 + a_2^2 + a_3^2$. It follows that A^{-1} , defined to be $|A|^{-2}\bar{A}$ is the inverse of A in the sense that $A \times A^{-1} = A^{-1} \times A = 1 =$

$(1, 0, 0, 0)$. Thus all the operations of arithmetic make sense on quaternions. One can add subtract, multiply, and divide. However, multiplying or dividing on the left is in general different from multiplying or dividing on the right. In general, the quaternions can be thought of as a number system containing a copy of the real numbers, infinitely many copies of the complex numbers, and a copy of ordinary three-dimensional vector space. That kind of generality causes a few restrictions in what can be said about quaternions.

Show that in quaternions the equation $X^2 + r^2 = 0$, where r is a positive real number, identified with the quaternion $R = (r, 0, 0, 0)$, is satisfied precisely by the quaternions $X = x + \xi$ such that $x = 0$, $|\xi| = r$, that is, by all the points on the sphere of radius r about the origin as center in three-dimensional space. In other words, in quaternions the square roots of negative real numbers are simply the nonzero vectors in three-dimensional space. Thus, even though quaternions act “almost” like the complex numbers, the fact that multiplication is not commutative makes a great difference when polynomial algebra is considered. In the quaternions, just as in the complex numbers, a linear equation can have only one solution. In contrast, while a quadratic equation can have only two solutions in the complex numbers, such an equation may have an uncountable infinity of solutions in the quaternions. Where does the proof that the number of roots does not exceed the degree of the equation, given above (Problem 37.2), break down? You can show nevertheless that if $R = (r, 0, 0, 0)$ is a *real* solution of the equation $P(X) = 0$, then $X - r$ does divide $P(X)$, and so a quaternion polynomial cannot have a number of *real* solutions that exceeds its degree. This is because r commutes with every quaternion.

- 37.8.** The results of Abel and Galois represent the outcome of many centuries of attempts to express the roots of a polynomial algebraically in terms of its coefficients. Meanwhile, computing the roots numerically from the coefficients—the approach followed in China and Japan—was a problem that had been solved centuries earlier. What are the advantages and disadvantages of each approach to the problem?
- 37.9.** Consider the set of all real and complex numbers that can be located in the complex plane using only straight lines and circles (straightedge and compass) starting from two arbitrary points labeled 0 and 1. Denote this set \mathbf{E} . We could call these “Euclidean numbers” or “constructible numbers” if we wished to invent a name for them. It is not difficult to show that if the complex numbers a , b , and c belong to \mathbf{E} , then so do the roots of the quadratic equation $ax^2 + bx + c = 0$. Thus the Euclidean numbers are “quadratically closed.” Are they also *algebraically* closed, in the sense that the roots of *every* polynomial with coefficients in \mathbf{E} also lie in \mathbf{E} ? [*Hint*: Consider the problem of duplicating the cube. What number must be constructed to solve this problem, and what equation does it satisfy?]

Projective and Algebraic Geometry and Topology

No area of modern mathematics is so difficult to summarize as geometry. Besides the projective and descriptive geometry developed during the Renaissance in connection with painting and engineering, the ancient problem of the parallel postulate resurfaced in the eighteenth century, leading to non-Euclidean geometry in the nineteenth. Descartes' analytic geometry infused the subject with algebra and, when algebra had mastered complex numbers, gave rise to the subject of algebraic geometry. Meanwhile, the application of calculus brought about the creation of differential geometry. The complicated curves and surfaces that could be handled by means of algebra and calculus led to further generalizations, and combinatorial, algebraic, and differential topology were the result. Finally, as an underpinning for all these subjects, the infusion of set theory into geometry brought about point-set topology in the twentieth century. We must employ drastic principles of selection to give even a hint of understanding of all this profound and beautiful mathematics. When we last looked at geometry, in Chapter 31, we discussed projective and descriptive geometry up to the mid-seventeenth century. In the present chapter, we shall discuss a small sampling of later developments in projective and algebraic geometry, along with various types of the more general geometry known as topology.

38.1. PROJECTIVE GEOMETRY

Projective geometry underwent a rapid development in the two centuries from 1650 to 1850. We shall discuss a sampling of the results and principles discovered.

38.1.1. Newton's Degree-Preserving Mapping

Newton described the mapping shown in Fig. 38.1 (Whiteside, 1967, Vol. VI, p. 269), in which the parallel lines BL and AO , the points A , B , and O are fixed from the outset, and the angle θ are specified in advance. These parameters determine the distances h and Δ and the angle φ . To map the figure GHI to its image ghi , first project each point G parallel to BL so as to meet the extension of AB at a point D . Next, draw the line OD meeting BL in point d . Finally, from d along the line making angle θ with BL , choose the image point

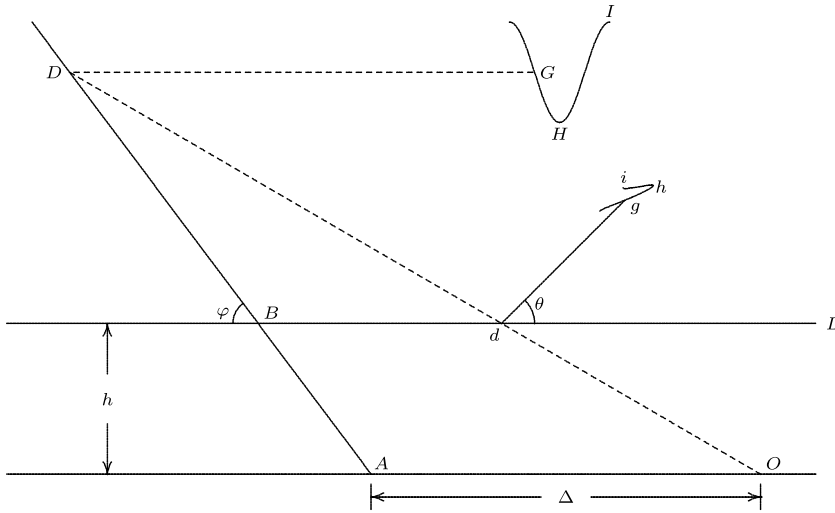


Figure 38.1. Newton's degree-preserving projection.

g so that $gd : Od :: GD : OD$. The original point, according to Newton, had coordinates (BD, DG) and its image the coordinates (Bd, dg) . If we let $x = BD$, $y = DG$, $\xi = Bd$, and $\eta = dg$, the coordinate transformations in the two directions are

$$(x, y) \mapsto (\xi, \eta) = \left(\frac{\Delta x \sin \varphi}{h + x \sin \varphi}, \frac{hy}{h + x \sin \varphi} \right),$$

$$(\xi, \eta) \mapsto \left(\frac{h\xi}{(\Delta - \xi) \sin \varphi}, \frac{\Delta\eta}{\Delta - \xi} \right).$$

Newton noted that this kind of projection preserves the degree of an equation. Hence a conic section will remain a conic section, a cubic curve will remain a cubic curve, and so on, under such a mapping. In fact, if a polynomial equation $p(x, y) = 0$ is given whose highest-degree term is $x^m y^n$, then every term $x^p y^q$, when expressed in terms of ξ and η , will be a multiple of $\xi^p \eta^q / (\Delta - \xi)^{p+q}$, so that if the entire equation is converted to the new coordinates and then multiplied by $(\Delta - \xi)^{m+n}$, this term will become $\xi^p \eta^q (\Delta - \xi)^{m+n-p-q}$, which will be of degree $m + n$. Thus the degree of an equation does not change under Newton's mapping. These mappings are special cases of the transformations known as *fractional-linear* or *Möbius* transformations, after August Ferdinand Möbius (1790–1868), who developed them more fully. They play a vital role in algebraic geometry and complex analysis, being the only one-to-one analytic mappings of the extended complex plane onto itself. According to Coolidge (1940, p. 269), Edward Waring remarked in 1762 that fractional-linear transformations were the most general degree-preserving transformations.

38.1.2. Brianchon

Pascal's work on the projective properties of conics was extended by Charles Julien Brianchon (1785–1864), who was also only a teenager when he proved what is now recognized as the dual of Pascal's theorem that the pairs of opposite sides of a hexagon inscribed in a conic meet in three collinear points. Pascal's theorem in the case of a circle is illustrated

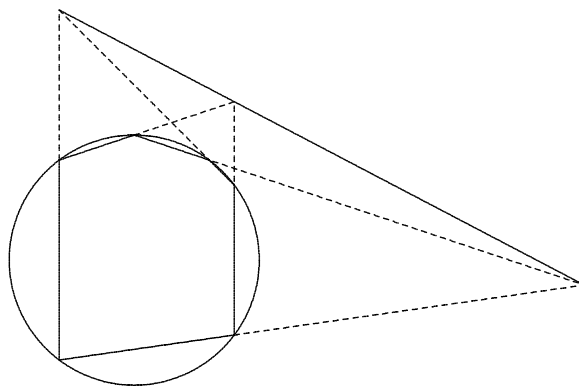


Figure 38.2. Pascal's theorem for a circle.

in Fig. 38.2. Brianchon's theorem asserts that the diagonals of a hexagon circumscribed about a conic are concurrent. The case of an ellipse is illustrated in Fig. 38.3.

38.1.3. Monge and his School

After a century of relative neglect, projective geometry revived at the Ecole Polytechnique under the students of Gaspard Monge (1746–1818), who was a master of the application of calculus to geometry. Felix Klein (1926, pp. 77–78) described his school as distinguished by “the liveliest spatial intuition combined in the most natural way possible with analytic operations.” Klein went on to say that Monge taught his students to make physical models, “not to make up for the deficiencies of their intuition but to develop an already clear and lively intuition.” As a military engineer, Monge had used his knowledge of geometry to design fortifications. His work in this area was highly esteemed by his superiors and declared a military secret. He wrote a book on descriptive geometry and one on the applications of analysis to geometry, whose influence appeared in the work of his students. Klein says of the second book that it “reads like a novel.” In this book, Monge analyzed quadric surfaces with extreme thoroughness.

Monge is regarded as the founder of descriptive geometry, which is based on the same principles of perspective as projective geometry but more concerned with the mechanics

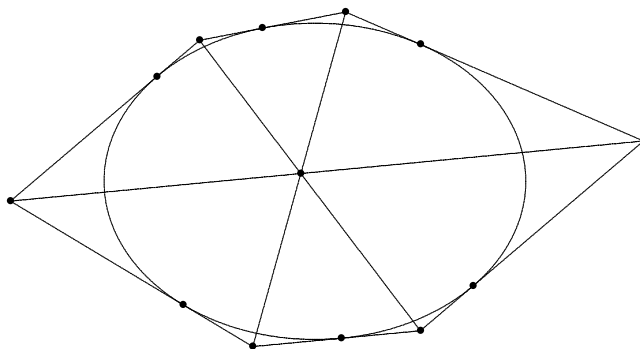


Figure 38.3. Brianchon's theorem for an ellipse.

of representing three-dimensional objects properly in two dimensions and the principles of interpreting such representations. Monge himself described the subject as the science of giving a complete description in two dimensions of those three-dimensional objects that can be defined geometrically. As such, it continues to be taught today under other names, such as mechanical drawing; it is the most useful form of geometry for engineers.

Monge's greatest student (according to Klein) was Jean-Victor Poncelet (1788–1867). He participated as a military engineer in Napoleon's invasion of Russia in 1812, was wounded, and spent a year in a Russian prison, where he busied himself with what he had learned from Monge. Returning to France, he published his *Treatise on the Projective Properties of Figures* in 1822, the founding document of modern projective geometry. Its connection with the work of Desargues shows in the first chapter, where Poncelet says he will be using the word *projective* in the same sense as the word *perspective*. In Chapter 3 he introduces the idea that all points at infinity in a plane can be regarded as belonging to a single line at infinity.¹ These concepts brought out fully the duality between points and lines in a plane and between points and planes in three-dimensional space, so that interchanging these words in a theorem of projective geometry results in another theorem. The theorems of Pascal and Brianchon, as mentioned, are dual to each other.

38.1.4. Steiner

The increasing algebraization of geometry was opposed by the Swiss mathematician Jacob Steiner (1796–1863), described by Klein (1926, pp. 126–127) as “the only example known to me. . . of the development of mathematical abilities after maturity.” Steiner had been a farmer up to the age of 17, when he entered the school of the Swiss educational reformer Johann Heinrich Pestalozzi (1746–1827), whose influence was widespread, extending through the philosopher/psychologist Johann Friedrich Herbart (1776–1841), a mathematically inclined philosopher whose attempts to quantify sense impressions was an early form of mathematical psychology,² down to Riemann, as will be explained in the next section.

Steiner was a peculiar character in the history of mathematics, who when his own creativity was in decline, adopted the ideas of others as his own without acknowledgment (see Klein, 1926, p. 128). But in his best years, around 1830, he had the brilliant idea of building space using higher-dimensional objects such as lines and planes instead of points, recognizing that these objects were projectively invariant. He sought to restore the ancient Greek “synthetic” approach to geometry, the one we have called *metric-free*, that is, independent of numbers and the concept of length. To this end, in his 1832 work on geometric figures he considered a family of mappings of one plane on another that resembles Newton's projection. Klein (1926, p. 129) found nothing materially new in this work, but admired the systematization that it contained. The Steiner principle of successively building more and more intricate figures by allowing simpler ones to combine geometrically was novel and had its uses, but according to Klein, encompassed only one part of geometry.

¹Field and Gray (1987, p. 185) point out that Kepler had introduced points at infinity in a 1604 work on conic sections, so that a parabola would have two foci.

²Herbart's 1824 book *Psychologie als Wissenschaft, neu gegründet auf Erfahrung, Metaphysik, und Mathematik* (*Psychology as Science on a New Foundation of Experiment, Metaphysics, and Mathematics*) is full of mathematical formulas involving the strength of sense impressions, manipulated by the rules of algebra and calculus. Klein (1926, pp. 127–128) has nothing good to say about the more extreme recommendations of these men, calling their recommendations “pedagogical monstrosities.”

38.1.5. Möbius

Projective geometry was enhanced through the barycentric calculus invented by Möbius and expounded in a long treatise in 1827. This work contained a number of very useful innovations. Möbius' use of barycentric coordinates to specify the location of a point anticipated vector methods by some 20 years and proved its value in many parts of geometry. He used his barycentric coordinates to classify plane figures in new ways. As he explained in Chapter 3 of the second section of his barycentric calculus (Baltzer, 1885, pp. 177–194), if the vertices of a triangle were specified as A, B, C , and one considered all the points that could be written as $aA + bB + cC$, with the lengths of the sides and the proportions of the coefficients $a : b : c$ given, all such figures would be congruent (he used the phrase “equal and similar”). If one specified only the proportions of the sides instead of their lengths, all such figures would be similar. If one specified only the proportions of the coefficients, the figures would be in an *affine* relationship, a word still used to denote a linear transformation followed by a translation in a vector space. Finally, he introduced the relation of equality (in area).

The work of Möbius on the barycentric calculus was reviewed by a French author with the initials A.C. It has been thought that this was Cauchy,³ Grattan-Guinness says, however, that the reviewer was probably not Cauchy. The reviewer, as reported by Baltzer (1885, pp. xi–xii), was cautious at first, saying that the work was “a different method of analytic geometry whose foundation is certainly not so simple; only a deeper study can enable us to determine whether the advantages of this method will repay the difficulties.” After reporting on the new classification of figures in Part 2, he commented:

One must be very confident of taking a large step forward in science to burden it with so much new terminology and to demand that your readers follow you in investigations presented to them in such a strange manner.

Finally, after reporting some of the results from Part 3, he concluded that, “It seems that the author of the barycentric calculus is not familiar with the general theory of duality between the properties of systems of points and lines established by M. Gergonne.”⁴ This duality (*gegenseitiges Entsprechen*) was discussed in Chapters 4 and 5 of Part 3. This comment apparently alludes to ideas in Gergonne's papers that the reviewer found missing in the work of Möbius. Chapters 4 and 5 contain some of the most interesting results in the work. Chapter 4, for example, discusses conic sections and uses the barycentric calculus to prove that two distinct parabolas can be drawn through four coplanar points, provided none of them lies inside the triangle formed by the other three.

Möbius is best remembered for two concepts, the Möbius transformation, and the Möbius band. A Möbius transformation, by which we now understand a mapping of the complex plane into itself, $z \mapsto w$, of the form

$$w = \frac{az + b}{cz + d}, \quad ad - bc \neq 0,$$

³Cauchy was able to read German, not a common accomplishment for French mathematicians in the 1820s, when the vast majority of mathematical papers of significance were written in French.

⁴Joseph Gergonne (1771–1859). Besides his work in geometry, he is best remembered as the founder (in 1810) of the journal *Annales des mathématiques pures et appliquées*.

can be found in his 1829 paper on metric relations in line geometry. He gave such transformations with real coefficients in terms of the two coordinates (x, y) , the real and imaginary parts of what we now write as the complex number z , and showed that they were the most general one-to-one transformations that preserve collinearity. The Möbius band is discussed in Section 38.3 below.

38.2. ALGEBRAIC GEOMETRY

Like Descartes, Newton made a classification of curves according to the degree of the equations that represent them or, rather, according to the maximal number of points in which they could intersect a straight line. As Descartes had argued for the use of any curves that could be generated by one parameter, excluding spirals and the quadratrix because they required two independent motions to be coordinated, Newton likewise argued that geometers should either confine themselves to conic sections or else allow any curve having a clear description. In his *Universal Arithmetick*, he mentioned in particular the trochoid,⁵ which makes it possible to divide an angle into any number of equal parts, as a useful curve that is simple to describe.

Descartes had begun the subject of algebraic geometry by classifying algebraic curves into “genera,” and, as just shown, Newton gave an alternative classification of curves, also based on algebra, although he included some curves that we would call transcendental, which could intersect a line in infinitely many points. The general study of algebraic curves $p(x, y) = 0$, where $p(x, y)$ is a polynomial in two variables, began with Colin Maclaurin (1698–1746), who in his *Geometria organica* of 1720 remarked that a cubic curve was not uniquely determined by nine points, even though nine points apparently suffice to determine the coefficients of any polynomial $p(x, y)$ of degree 3, up to proportionality and hence ought to determine a unique curve $p(x, y) = 0$. Two *distinct* cubic curves generally intersect in nine points, so that *some* sets of nine points do not determine the curve uniquely. This fact was later (1748) noted by Euler as well, and finally, by Gabriel Cramér (1704–1752), who also noted Maclaurin’s priority in the discovery that a curve of degree m and a curve of degree n meet generally in mn points.

This curious fact is called *Cramér’s paradox* after Cramér published it in a 1750 textbook on algebraic curves. Although he correctly explained why more than one curve of degree n can sometimes be made to pass through $n(n + 3)/2$ points—because the equations for determining the coefficients from the coordinates of the points might not be independent—he noted that in that case there were actually infinitely many such curves. That, he said, was a real paradox. Incidentally, it was in connection with the determination of the coefficients of an algebraic curve through given points that Cramér stated Cramér’s rule for solving a system of linear equations by determinants.⁶

⁵A trochoid is the locus of a point rigidly attached to a rolling wheel. If the point lies between the rim and the center, the trochoid is called a *curtate cycloid*. If the point lies outside the rim, the trochoid is a *prolate cycloid*. If the point is on the rim, the trochoid is called simply a *cycloid*. The names come from the Greek words *trokhós* (*wheel*) and *kýklos* (*circle*).

⁶As mentioned in Chapter 24, the solution of linear equations by determinants had been known to Seki Kōwa and Leibniz. Thus, Cramér has two mathematical concepts named after him, and in both cases he was the third person to make the discovery.

38.2.1. Plücker

A number of excellent German, Swiss, and Italian geometers arose in the nineteenth century. As an example, we take Julius Plücker (1801–1868), who was a professor at the University of Bonn for the last 30 years of his life. Plücker himself remembered (Coolidge, 1940, p. 144) that when young he had discovered a theorem in Euclidean geometry: The three lines containing the common chords of pairs of three intersecting circles are all concurrent. Plücker's proof of this theorem is simplicity itself. Suppose that the equations of the three circles are $A = 0$, $B = 0$, $C = 0$, where each equation contains $x^2 + y^2$ plus linear terms. By subtracting these equations in pairs, we get the quadratic terms to drop out, leaving the equations of the three lines containing the three common chords: $A - B = 0$, $A - C = 0$, $B - C = 0$. But it is manifest that any two of these equations imply the third, so that the point of intersection of any two also lies on the third line.

Plücker's student Felix Klein (1926, p. 122) described a more sophisticated specimen of this same kind of reasoning by Plücker to prove Pascal's theorem that the opposite sides of a hexagon inscribed in a conic, when extended, intersect in three collinear points. The proof goes as follows: The problem involves two sets, each containing three lines, six of whose nine pairwise intersections lie on a conic section. The conic section has an equation of the form $q(x, y) = 0$, where $q(x, y)$ is quadratic in both x and y . Represent each line by a linear polynomial of the form $a_jx + b_jy + c_j$, the j th line being the set of (x, y) where this polynomial equals zero, and assume that the lines are numbered in clockwise order around the hexagon. Form the polynomial

$$s(x, y) = (a_1x + b_1y + c_1)(a_3x + b_3y + c_3)(a_5x + b_5y + c_5) \\ - \mu(a_2x + b_2y + c_2)(a_4x + b_4y + c_4)(a_6x + b_6y + c_6)$$

with the parameter μ to be chosen later. This polynomial vanishes at all nine intersections of the lines. Line 1, for example, meets lines 2 and 6 inside the conic and line 4 outside it.

Now, when y is eliminated from the equations $q(x, y) = 0$ and $s(x, y) = 0$, the result is an equation $t(x) = 0$, where $t(x)$ is a polynomial of degree at most 6 in x . This polynomial must vanish at all of the simultaneous zeros of $q(x, y)$ and $s(x, y)$. We know that there are six such zeros for every μ . However, it is very easy to choose μ so that there will be a seventh common zero. With that choice of μ , the polynomial $t(x)$ must have seven zeros, and hence must vanish identically. But since $t(x)$ was the result of eliminating y between the two equations $q(x, y) = 0$ and $s(x, y) = 0$, it now follows that $q(x, y)$ divides $s(x, y)$. That is, the equation $s(x, y) = 0$ can be written as $(ax + by + c)q(x, y) = 0$. Hence its solution set consists of the conic and the line $ax + by + c = 0$, and this line must contain the other three points of intersection.

Conic sections and quadratic functions in general continued to be a source of new ideas for geometers during the early nineteenth century. Plücker liked to use homogeneous coordinates to give a symmetric description of a quadric surface (a surface whose equation is $p(x, y, z) = 0$, where $p(x, y, z)$ is a polynomial of degree 2). To take the simplest example, consider the sphere of radius 2 in three-dimensional space with center at $(2, 3, 1)$, whose equation is

$$(x - 2)^2 + (y - 3)^2 + (z - 1)^2 = 4.$$

If x , y , and z , are replaced by ξ/τ , η/τ , and ζ/τ and each term is multiplied by τ^2 , this equation becomes a homogeneous quadratic relation in the four variables (ξ, η, ζ, τ) :

$$(\xi - 2\tau)^2 + (\eta - 3\tau)^2 + (\zeta - \tau)^2 = 4\tau^2.$$

The sphere of unit radius centered at the origin then has the simple equation $\tau^2 - \xi^2 - \eta^2 - \zeta^2 = 0$. Plücker introduced homogeneous coordinates in 1830. One of their advantages is that if $\tau = 0$, but the other three coordinates are not all zero, the point (ξ, η, ζ, τ) can be considered to be located on a sphere of infinite radius. The point $(0, 0, 0, 0)$ is excluded, since it seems to correspond to all points at once.

Homogeneous coordinates correspond very well to the ideas of projective geometry, in which a point in a plane is identified with all the points in three-dimensional space that project to that point from a point outside the plane. If, for example, we take the center of projection as $(0, 0, 0)$ and identify the plane with the plane $z = 1$, that is, each point (x, y) is identified with the point $(x, y, 1)$, the points that project to (x, y) are all points (tx, ty, t) , where $t \neq 0$. Since the equation of a line in the (x, y) -plane has the form $ax + by + c = 0$, one can think of the coordinates (a, b, c) as the coordinates of the line. Here again, multiplication by a nonzero constant does not affect the equation, so that these coordinates can be identified with (ta, tb, tc) for any $t \neq 0$. Notice that the condition for the point (x, y) to lie on the line (a, b, c) is that $\langle (a, b, c), (x, y, 1) \rangle = a \cdot x + b \cdot y + c \cdot 1 = 0$, and this condition is unaffected by multiplication by a constant. The duality between points and lines in a plane is then clear. Any triple of numbers, not all zero, can represent either a point or a line, and the incidence relation between a point and a line is symmetric in the two. We might as well say that the line lies on the point as that the point lies on the line.

Equations can be written in either line coordinates or point coordinates. For example, the equation of an ellipse can be written in homogeneous point coordinates (ξ, η, ζ) as

$$b^2c^2\xi^2 + a^2c^2\eta^2 = a^2b^2\zeta^2,$$

or in line coordinates (λ, μ, ν) as

$$a^2\lambda^2 + b^2\mu^2 = c^2\nu^2,$$

where the geometric meaning of this last expression is that the line (λ, μ, ν) is tangent to the ellipse.

38.2.2. Cayley

Homogeneous coordinates provided important invariants and covariants⁷ in projective geometry. One such invariant under orthogonal transformations (those that leave the sphere

⁷According to Klein (1926, p. 148), the distinction between an invariant and a covariant is not essential. An algebraic expression that remains unchanged under a family of changes of coordinates is a covariant if it contains variables, and it is an invariant if it contains only constants.

fixed) is the dihedral angle between two planes $Ax + By + Cz = D$ and $A'x + B'y + C'z = D'$, given by

$$\arccos \left(\frac{AA' + BB' + CC'}{\sqrt{A^2 + B^2 + C^2} \sqrt{(A')^2 + (B')^2 + (C')^2}} \right). \quad (38.1)$$

In his “Sixth memoir on quantics,” published in the *Transactions of the London Philosophical Society* in 1858, Cayley fixed a “quantic” (quadratic form) $\sum \alpha_{ij} u_i u_j$, whose zero set was a quadric surface that he called the *absolute*, and defined angles by analogy with Eq. (38.1) and other metric concepts by a similar analogy. In this way he obtained the *general projective metric*, commonly called the *Cayley metric*. It allowed metric geometry to be included in projective geometry. As Cayley said, “Metrical geometry is thus a part of descriptive geometry and descriptive geometry is all geometry.” By suitable choices of the absolute, one could obtain the geometry of all kinds of quadric curves and surfaces, including the non-Euclidean geometries studied by Gauss, Lobachevskii, Bólyai, and Riemann, all of which will be discussed in Chapter 40. Klein (1926, p. 150) remarked that Cayley’s models were the most convincing proof that these geometries were consistent.

38.3. TOPOLOGY

Projections distort the shape of geometric objects, so that some metric properties are lost. Some properties, however, remain because the number of intersections of two lines does not change. The study of space focusing on such very general properties as connections and intersections has been known by various names over the centuries. Latin has two words, *locus* and *situs*, meaning roughly *place* and *position*. The word *locus* is one that we still use today to denote the path followed by a point moving subject to stated constraints, although, since the introduction of set theory, a locus is more often thought of statically as the set of points satisfying a given condition. It was the translation of the Greek word *tópos* used by Pappus for the same concept. Since *locus* was already in use, Leibniz fastened on *situs* and mentioned the need for a geometry or analysis of *situs* in a 1679 letter to Huygens.⁸ The meaning of *geometria situs* and *analysis situs* evolved gradually. It seems to have been Johann Benedict Listing (1808–1882) who, some time during the 1830s, realized that the Greek root was available. The word *topology* first appeared in the title of his 1848 book *Vorstudien zur Topologie (Prolegomena to Topology)*. Like geometry itself, topology has bifurcated several times, so that one can now distinguish combinatorial, algebraic, differential, and point-set topology.

38.3.1. Combinatorial Topology

The earliest result that deals with the combinatorial properties of figures is now known as the *Euler characteristic*, although Descartes is entitled to some of the credit. In a work on polyhedra that he never published, Descartes defined the *solid angle* at a vertex of a

⁸This letter was published in the 1888 edition of Huygens’ *Œuvres complètes*, Vol. 8, p. 216. From the context it appears that Leibniz was calling for some simple way of expressing position “as algebra expresses magnitude.” If so, perhaps we now have what he wanted, in the form of vector analysis.

closed polyhedron to be the difference between a complete revolution (4 right angles) and the sum of the angles at that vertex. He noted that the sum of the solid angles in any closed polyhedron was exactly eight right angles. Descartes' work was found among his effects after he died. By chance, Leibniz saw it a few decades later and made a copy of it. When it was found among Leibniz' papers, it was finally published. In the eighteenth century, Euler discovered that the sum of the angles at the vertices of a closed polyhedron was $4V - 8$ right angles, where V is the number of vertices. Euler noted the equivalent fact that the number of faces and vertices exceeded the number of edges by 2. That is the formula now generally called *Euler's formula*:

$$V - E + F = 2.$$

Somewhat peripheral to the general subject of topology was Euler's analysis of the famous problem of the seven bridges of Königsberg⁹ in 1736. In Euler's day there were two islands in the middle of the River Pregel, which flows through Königsberg (now Kaliningrad, Russia). These islands were connected to each other by a bridge, and one of them was connected by two bridges to each shore, the other by one bridge to each shore. The problem was to go for a walk and cross each bridge exactly once, returning, if possible, to the starting point. In fact, as one can easily see, it is impossible even to cross each bridge exactly once without boating or swimming across the river. Returning to the starting point merely adds another condition to a condition that is already impossible to fulfill. Euler proved this fact by labeling the two shores and the two islands A , B , C , and D and representing a hypothetical stroll as a "word," such as $ABCBD$, in which the bridges are "between" the letters. He showed that any such path as required would have to be represented by an 8-letter word containing three of the letters twice and the other letter three times, which is obviously impossible. This topic belongs to what is now called graph theory; it is an example of the problem of unicursal tracing.

38.3.2. Riemann

The study of analytic functions of a complex variable turned out to require some concepts from topology. These issues were touched on in Riemann's 1851 doctoral dissertation at Göttingen, "Grundlagen für eine allgemeine Theorie der Functionen einer veränderlichen complexen Grösse" ("Foundations for a general theory of functions of a complex variable"). Although all analytic functions of a complex variable, both algebraic and transcendental, were encompassed in Riemann's ideas, he was particularly interested in algebraic functions—that is, functions $w = f(z)$ that satisfy a nontrivial polynomial equation $p(z, w) = 0$. Algebraic functions are essentially and unavoidably multivalued. To take the simplest example, where $z - w^2 = 0$, every complex number $z = a + bi$ has two distinct complex square roots:

$$w = \pm(u + iv), \quad \text{where } u = \sqrt{\frac{\sqrt{a^2 + b^2} + a}{2}} \quad \text{and} \quad v = \operatorname{sgn}(b) \sqrt{\frac{\sqrt{a^2 + b^2} - a}{2}}.$$

⁹This problem, because it is simple to state and has an elegant solution, is extremely popular in mathematics survey courses for nonmajors. However, its significance in the wider world of topology is quite small, so that it contributes to a generally distorted popular picture of what mathematicians have achieved.

The square roots of the non-negative real numbers that occur here are assumed nonnegative. There is no way of choosing just one of the two values at each point that will result in a continuous function $w = \sqrt{z}$. In particular, it is easy to show that any such choice must have a discontinuity at some point of the circle $|z| = 1$.

One way to handle this multivaluedness was to take two copies of the z -plane, labeled with subscripts as z_1 and z_2 , and place one of the square roots in one plane and the other in the other. This technique was used by Cauchy and had been developed into a useful way of looking at complex functions by Victor Puiseux (1820–1883) in 1850. Indeed, Puiseux seems to have had the essential insight—preventing z from making a complete circuit around 0—that can be found in Riemann’s work, although differently expressed. Riemann is known to have seen the work of Puiseux, although he did not cite it in his own work. He generally preferred to work out his own way of doing things and tended to ignore earlier work by other people. The difficulty with choosing one square root and sticking to it is that a single choice cannot be continuous on a closed path that encloses the origin but does not pass through it. Somewhere on such a path, there will be nearby points at which the function assumes two values that are close to being negatives of each other.

Riemann had the idea of cutting the two copies of the z -plane along a line running from zero to infinity (both being places where there is only one square root, assuming a bit about complex infinity). These two points are called *branch points*. Then if the lower edge of each plane is imagined as being glued to the upper edge of the other,¹⁰ the result is a single connected surface in which the origin belongs to both planes. On this new surface a continuous square-root function can be defined. It was the gluing that was really new here. Cauchy and Puiseux both had the idea of cutting the plane to keep a path from winding around a branch point and of using different copies of the plane to map different branches of the function.

Riemann introduced the idea of a *simply connected surface*, one that is disconnected by any cut from one boundary point to another that passes through its interior without intersecting itself. He stated as a theorem that the result of such a cut would be two simply connected surfaces. In general, when a connected surface is cut by a succession of such *crosscuts*, as he called them, the number of crosscuts minus the number of connected components that they produce, plus 1, is a constant, called the *order of connectivity* of the surface. A sphere, for example, can be thought of as a square with adjacent edges glued together, as in Fig. 38.3. It is simply connected (has order of connectivity 0) because a diagonal cut divides it into two components. The torus, on the other hand, can be thought of as a square with opposite edges identified (see Fig. 38.4). To separate this surface into two components, it is necessary to cut the square at least twice, for example, either along both diagonals or through its center along two lines parallel to the sides. No single cut will do. The torus is thus doubly connected, with order of connectivity 1).

38.3.3. Möbius

One fact that had been thought well established about polyhedra was that in any polyhedron it was possible to direct the edges in such a way that one could trace around the boundary of each face by following the prescribed direction of its edges. Each face would be always to the left or always to the right as one followed the edges around it while looking at it from

¹⁰This gluing is shown in more detail in Chapter 41.

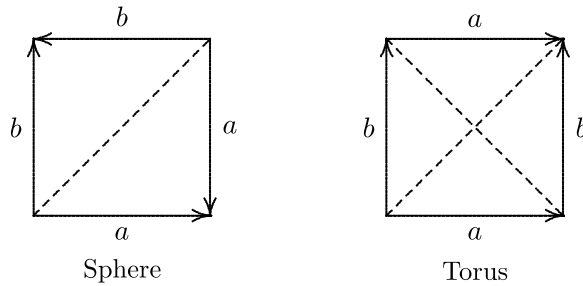


Figure 38.4. Left: The sphere, regarded as a square with edges identified, is disconnected by a diagonal cut. Right: Two cuts are required to disconnect the torus.

outside the polyhedron. This fact was referred to as the *edge law* (*Gesetz der Kanten*). The first discovery of a polyhedron that violated this condition¹¹ was due to Möbius, sometime during the late 1850s. Möbius did not publish this work, although he did submit some of it to the Paris Academy as his entry to a prize competition in 1858. This work was edited and introduced by Curt Reinhardt (dates unknown) and published in Vol. 2 of Möbius’ collected works. There in the first section, under the heading “one-sided polyhedra,” is a description of the Möbius band as we now know it (Fig. 38.5). After describing it, Möbius went on to say that although a triangulated polyhedron whose surface was two-sided will apparently contain only two-sided bands, *nevertheless a triangulated one-sided polyhedron can contain both one- and two-sided bands.*

Möbius explored polyhedra and made a classification of them according to the number of boundary curves they possessed. He showed how more complicated polyhedra could be produced by gluing together a certain set of basic figures. He found an example of a triangulated polyhedron consisting of 10 triangles, six vertices, and 15 edges, rather than 14, as would be expected from Euler’s formula for a closed polyhedron: $V - E + F = 2$. This figure is the projective plane, and cannot be embedded in three-dimensional space. If one of the triangular faces is removed, the resulting figure is the Möbius band, which can be embedded in three-dimensional space.

38.3.4. Poincaré’s *Analysis situs*

Henri Poincaré (1854–1912) dealt with topological considerations frequently in his work in both complex function theory and differential equations. To set everything that he discovered down in good order, he wrote a treatise on topology called *Analysis situs* in 1895, published in the *Journal de l’Ecole Polytechnique*. This paper has been regarded as the founding document of modern algebraic topology.¹² He introduced the notion of homologous curves—curves that (taken together) form the boundary of a surface. This notion could be formalized, so that one could consider formal linear combinations (now called *chains*) $C = n_1C_1 + \dots + n_rC_r$ of oriented curves C_i with integer coefficients n_i . The interpretation of such a combination came from analysis: A line integral over C was interpreted as the

¹¹In fact, a closed nonorientable polyhedron cannot be embedded in three-dimensional space, so that the edge law is actually true for *closed* polyhedra in three-dimensional space.

¹²Poincaré followed this paper with a number of supplements over the next decade.

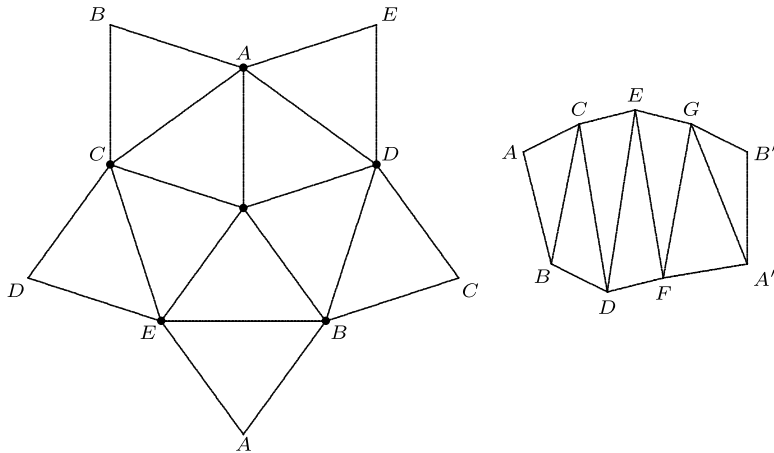


Figure 38.5. Left: the projective plane triangulated and cut open. If two opposite edges with corresponding endpoints are glued together, the figure becomes a Möbius band. In three-dimensional space it is not possible to glue all the edges together as indicated. Right: the Möbius band as originally described by Möbius.

number $I = n_1 I_1 + \dots + n_r I_r$, where I_j was the line integral over C_j . When generalized to k -dimensional manifolds (called *variétés* by Poincaré) and combined with the concept of the boundary of an oriented manifold as a cycle, this idea was the foundation of homology theory: The k -cycles (k -chains whose boundaries are the zero $(k - 1)$ -chain—Poincaré called them *closed varieties*) form a group, of which the k -cycles that are the boundary of a $(k + 1)$ -cycle form a subgroup. When two homologous cycles (cycles whose difference is a boundary) are identified, the resulting classes of cycles form the k th *homology group*. For example, in the sphere shown in Fig. 38.4, the diagonal that is drawn forms a cycle. This cycle is the complete boundary of the upper and lower triangles in the figure, and it turns out that any cycle on the sphere is a boundary. The first homology group of the sphere is therefore trivial (consists of only one element). For the torus depicted in Fig. 38.4, a and b are each cycles, but neither is a boundary, nor is any cycle $ma + nb$. On the other hand, the cycle formed by adding either diagonal to $a + b$ is the boundary of the two triangles with these edges. Thus, the first homology group of the torus can be identified with the set of cycles $ma + nb$. Any other cycle will be homologous to one of these.

Poincaré also introduced a second concept that has been of immense value in analyzing manifolds. He had been led to algebraic topology partly by his work in differential equations. In that connection, he imagined functions satisfying a set of differential equations and being permuted as a point moved around a closed loop. He was thus led to consider formal sums of loops starting and ending at a given point, two loops being equivalent if tracing them successively left the functions invariant. The resulting set of permutations was what he called the *fundamental group* or *first homotopy group*.¹³ He cautioned that, despite appearances, the first homotopy group was not the same thing as the first homology group, since there was no base point involved in the homology group. Moreover, he noted, while the order in which the cycles in a chain were traversed was irrelevant, the fundamental group was not necessarily

¹³In informal terms, a homotopy is a continuous deformation of one curve into another.

commutative. He suggested redefining the term *simply connected* to mean having a trivial fundamental group. He gave examples to show that the homology groups do not determine the topological nature of a manifold, exhibiting three three-dimensional manifolds all having the same homology groups, but different fundamental groups and therefore not topologically the same (homeomorphic). He then asked a number of questions about fundamental groups, one of which has become famous. *Given two manifolds of the same number of dimensions having the same homotopy groups, are they homeomorphic?* Like Fermat's last theorem, this question has been attacked by many talented mathematicians. Many cases of it were settled, but not the important case of the sphere in three dimensions. For that case, many partial results were produced, and many proofs were proposed for a positive answer to the question, but until recently, all such proofs were found wanting. At last, in the first decade of the twentieth century, the Russian mathematician Grigorii Perelman (b. 1966) of the Steklov Institute in St. Petersburg gave a proof of the positive answer.¹⁴

38.3.5. Point-Set Topology

Topology is sometimes popularly defined as “rubber-sheet geometry,” in the sense that the concepts it introduces are invariant under moving and stretching, provided that no tearing takes place. In the kinds of combinatorial topology just discussed, those concepts usually involve numbers in some form or other—the number of independent cycles on a manifold, the Euler characteristic (the number $V - E + F$ when a surface is partitioned), and so forth. But there are also topological concepts not directly related to number. One of these concepts, that of a Riemann surface, was designed for the needs of algebraic geometry and complex analysis. A quite different kind of topology, known as point-set topology, arose in complex and real analysis, but was mostly applied in real analysis, where it forms a large part of the subject matter.

The simplest and most intuitive of these concepts is that of *connectedness* or *continuity*. This word denotes a deep intuitive idea that was the source of many paradoxes in ancient times, such as the paradoxes of Zeno. In fact, it is impossible to prove the fundamental theorem of algebra without this concept.¹⁵ For analysts, it was crucial to know that if a continuous function was negative at one point on a line and positive at another, it must assume the value zero at some point between the two points (the intermediate-value property). That property eventually supplanted earlier definitions of continuity, and the property now taken as the definition of continuity is designed to make this proposition true. The clarification of the ideas surrounding continuity occurred in the early part of the nineteenth century, as mentioned in Chapter 34. Once serious analysis of this concept was undertaken, it became clear that many intuitive assumptions about the connectedness of curves and surfaces had been made from the beginning of deductive geometry. These continuity considerations complicated the theory of functions of a real variable for some decades until adequate explanations were found. A good example of such problems is provided by Dedekind's

¹⁴For this achievement and a number of other brilliant papers, which Perelman chose to publish on the Word-Wide Web rather than in a refereed journal, he was showered with honors, including a number of professorships, a prize of one million dollars, and the Fields Medal, the most prestigious award in the mathematical profession. Perelman, however, has refused to accept any of these honors.

¹⁵Even the second of the four proofs that Gauss gave, which is generally regarded as a purely algebraic proof, required the assumption that an equation of odd degree with real coefficients has a real solution—a fact that relies on connectedness.

construction of the real numbers, which will be discussed in Chapter 42 and which he presented as a solution to the problem of defining what is meant by a continuum.

PROBLEMS AND QUESTIONS

Mathematical Problems

- 38.1.** How can Pascal's theorem for an ellipse be deduced from the special case of a circle? How do you interpret the situation when one pair of opposite sides is parallel?
- 38.2.** Fill in the details of Plücker's proof of Pascal's theorem, as follows: Suppose that the equation of the conic is $q(x, y) = y^2 + r_1(x)y + r_2(x) = 0$, where $r_1(x)$ is a linear polynomial and $r_2(x)$ is quadratic. Choose coordinate axes not parallel to any of the sides of the inscribed hexagon and such that the x -coordinates of all of the vertices of the hexagon will be different, and also choose the seventh point where $t(x) = 0$ to have x -coordinate different from those of the six vertices. Then suppose that the polynomial generated by the three lines is $s(x, y) = y^3 + t_1(x)y^2 + t_2(x)y + t_3(x) = 0$, where $t_j(x)$ is of degree j , $j = 1, 2, 3$. Then there are polynomials $u_j(x)$ of degree j , $j = 1, 2, 3$, such that

$$s(x, y) = q(x, y)(y - u_1(x)) + (u_2(x)y + u_3(x)).$$

We need to show that $u_2 \equiv 0$ and $u_3 \equiv 0$. At the seven points on the conic where both $q(x, y)$ and $s(x, y)$ vanish, it must also be true that $u_2(x)y + u_3(x) = 0$. Rewrite the equation $q(x, y) = 0$ at these seven points as

$$(u_2y)^2 + r_1u_2(u_2y) + u_2^2r_2 = 0$$

and observe that at these seven points $u_2y = -u_3$, so that the polynomial $u_3^2 - r_1u_2u_3 + u_2^2r_2$, which is of degree 6, has seven distinct zeros. It must therefore vanish identically, and that means that

$$(2u_3 - r_1u_2)^2 = u_2^2(r_1^2 - 4r_2).$$

This means that either u_2 is identically zero, which implies that u_3 also vanishes identically, or else u_2 divides u_3 . Prove that in the second case the conic must be a pair of lines, and give a separate argument in that case.

- 38.3.** Consider the general cubic equation

$$Ax^3 + Bx^2y + Cxy^2 + Dy^3 + Ex^2 + Fxy + Gy^2 + Hx + Iy + J = 0,$$

which has 10 coefficients. Show that if this equation is to hold for the 10 points $(1, 0)$, $(2, 0)$, $(3, 0)$, $(4, 0)$, $(0, 1)$, $(0, 2)$, $(0, 3)$, $(1, 1)$, $(2, 2)$, $(1, -1)$, all 10 coefficients A, \dots, J must be zero. In general, then, it is not possible to pass a curve of degree 3 through any 10 points in the plane. Use linear algebra to show that it is always possible to pass a curve of degree 3 through any nine points, and that the curve is generally unique.

On the other hand, two *different* curves of degree 3 generally intersect in 9 points, a result known as Bézout's theorem after Etienne Bézout (1730–1783), who stated it around 1758, although Maclaurin had stated it earlier. How does it happen that while nine points generally determine a *unique* cubic curve, *two distinct* cubic curves generally intersect in nine points? [*Hint*: Suppose that a set of eight points $\{(x_j, y_j) : j = 1, \dots, 8\}$ is given for which the system of equations for A, \dots, J has rank 8. Although the system of linear equations for the coefficients is generally of rank 9 if another point is adjoined to this set, there generally is a point (x_9, y_9) , namely the ninth point of intersection of two cubic curves through the other eight points, for which the rank will remain at 8.]

Historical Questions

- 38.4.** What was the unique feature of Steiner's approach to geometry that made it an advance over previous work?
- 38.5.** In what sense are the theorems of Pascal (Chapter 31) and Brianchon dual to each other?
- 38.6.** In what way did Cayley's concept of the absolute unify geometry?

Questions for Reflection

- 38.7.** Explain Cramér's paradox, and why it is not a contradiction.
- 38.8.** Why are Möbius transformations useful in algebraic geometry and complex analysis?
- 38.9.** What underlying harmony and unity is responsible for the fact that Pascal's theorem can be proved both projectively (from the case of a circle) and algebraically, as Plücker did it?

Differential Geometry

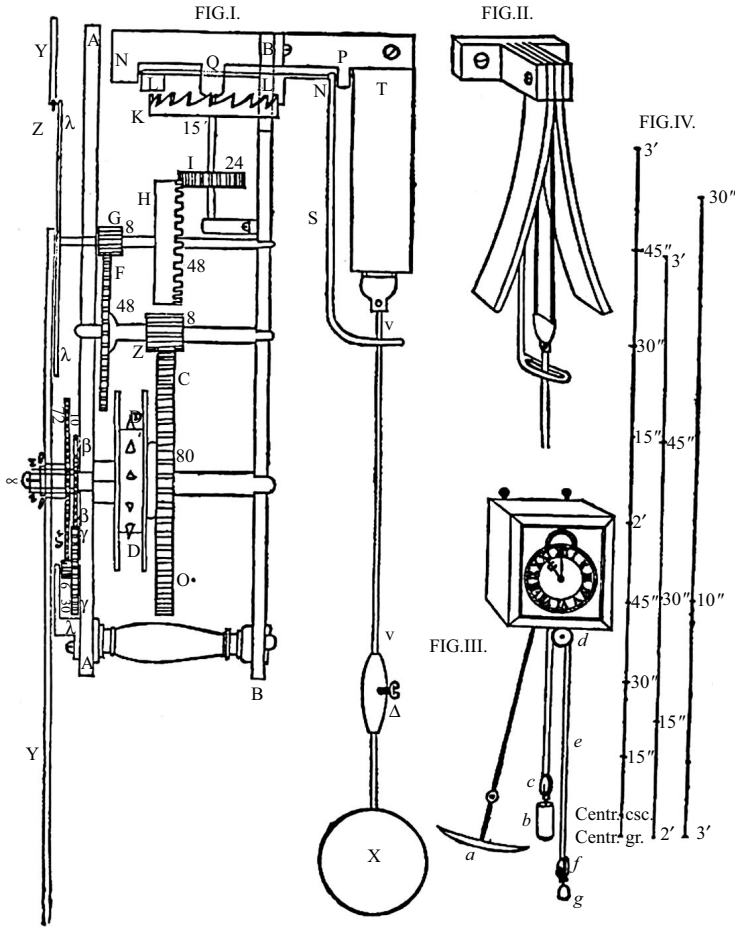
Differential geometry is the study of curves and surfaces (from 1852 on, manifolds) using the methods of differential calculus, such as derivatives and local series expansions. This history falls into natural periods defined by the primary subject matter: first, the tangents and curvatures of plane curves; second, the same properties for surfaces and curves in three-dimensional space; third, minimal surfaces and geodesics on surfaces; fourth, the application (conformal mapping) of surfaces on one another; fifth, extensions of all these topics to n -dimensional manifolds and global properties instead of local.

39.1. PLANE CURVES

Besides the study of tangent and normal lines to plane curves, which was begun in connection with analytic geometry, certain auxiliary curves were studied—in particular the involute, which is defined below. Measures of curvature, such as the osculating circle (the circle that fits a curve up to second order near a point) and the radius of curvature became a focus of attention. As mentioned in Chapter 34, Brook Taylor used the assumption that the restorative force on a stretched string is proportional to its curvature in order to study the vibrations of strings. This assumption showed very good intuition, since the curvature is the second derivative with respect to arc length; under the approximations used to get a linear model for this phenomenon from Newton's laws of motion, that assumption yields precisely the correct equation.

39.1.1. Huygens

Struik (1933) and Coolidge (1940, p. 319) agree that credit for the first exploration of secondary curves generated by a plane curve—the involute and evolute—occurred in Christiaan Huygens' work *Horologium oscillatorium (Of Pendulum Clocks)* in 1673, even though calculus had not yet been developed. The involute of a curve is the path followed by the endpoint of a taut string being wound onto the curve or unwound from it. Huygens did not give it a name; he simply called it the “line [curve] described by evolution.” There are as many involutes as there are points on the curve to begin or end the winding process.



A cycloidal pendulum clock, from Huygens' *Horologium oscillatorium*. Copyright © Stock Montage.

Huygens was seeking a truly synchronous pendulum clock, and he needed a pendulum that would have the same period of oscillation no matter how great the amplitude of the oscillation was.¹ Huygens found the mathematically ideal solution of the problem in two properties of the cycloid. First, a frictionless particle requires the same time to slide to the bottom of a cycloid no matter where it begins (the tautochrone property); second, the involute of a cycloid is another cycloid. He therefore designed a pendulum clock in which the pendulum bob was attached to a flexible leather strap that is confined between two inverted cycloidal arcs. The pendulum is thereby forced to fall along the involute of a cycloid and hence to trace another cycloid. Reality being more complicated than our

¹Despite the legend that Galileo observed a chandelier swinging and noticed that all its swings, whether wide or short, required the same amount of time to complete, that observation holds true for circular arcs only approximately and only for small amplitudes, as anyone who has done the experiment in high-school physics will have learned.

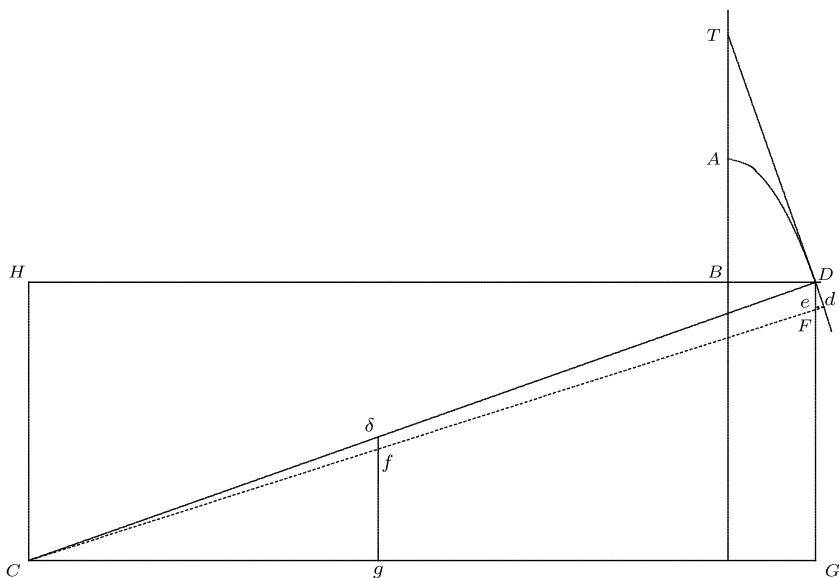


Figure 39.1. Newton's construction of the radius of curvature, from his posthumously published *Fluxions*.

dreams, however, this apparatus—like the mechanical drawing methods of Albrecht Dürer discussed in Chapter 31—does not really work any better than a circular pendulum.²

39.1.2. Newton

In his *Fluxions*, which was first published in 1736 after his death, even though it appears to have been written in 1671, Newton found the circle that best fits a curve. Struik (1933, 19, p. 99) doubted that this material was really in the 1671 manuscript. Be that as it may, the topic occurs as Problem 5 in the *Fluxions*: *At any given Point of a given Curve, to find the Quantity of Curvature.* Newton needed to find a circle tangent to the curve at a given point, which meant finding its center. However, Newton wanted not just any tangent circle. He assumed that if a circle was tangent to a curve at a point and “no other circle can be interscribed in the angles of contact near that point, . . . that circle will be of the same curvature as the curve is of, in that point of contact.” In this connection he introduced terms *center of curvature* and *radius of curvature* still used today. His construction is shown in Fig. 39.1, in which one unnecessary letter has been removed and the figure has been rotated through a right angle to make it fit the page. The weak point of Newton's argument was his claim that, “If *CD* be conceived to move, while it insists [remains] perpendicularly on the Curve, that point of it *C* (if you except the motion of approaching to or receding from the Point of Insistence *C*,) will be least moved, but will be as it were the Center of Motion.” Huygens had had this same problem with clarity. Where Huygens had referred to points

²The master's thesis of Robert W. Katsma at California State University at Sacramento in the year 2000 was entitled “An analysis of the failure of Huygens' cycloidal pendulum and the design and testing of a new cycloidal pendulum.” Katsma was granted patent 1992-08-18 in Walla Walla County for a cycloidal pendulum. However, the theoretical consensus is that such devices only decrease the accuracy of a good clock.

that *can be treated as* coincident, Newton used the phrase *will be as it were*. One can see why infinitesimal reasoning was subject to so much doubt at the time.

Newton also treated the problem of the cycloidal pendulum in his *Principia Mathematica*, published in 1687. Huygens had found the evolute of a complete arch of a cycloid. That is, the complete arch is the involute of the portion of two half-arches starting at the halfway point on the arch. In Proposition 50, Problem 33 of Book 1, Newton found the evolute for an arbitrary piece of the arch, which was a much more complicated problem. It was, however, once again a cycloid. This evolute made it possible to limit the oscillations of a cycloidal pendulum by putting a complete cycloidal frame in place to stop the pendulum when the thread was completely wound around the evolute.

39.1.3. Leibniz

Leibniz' contributions to differential geometry began in 1684, when he gave the rules for handling what we now call differentials. His notation is essentially the one we use today. He regarded x and $x + dx$ as infinitely near values of x and v and dv as the corresponding infinitely near values of v on a curve defined by an equation relating x and v . At a maximum or minimum point he noted that $dv = 0$, so that the equation defining the curve had a double root (v and $v + dv$) at that point. He noted that the two cases could be distinguished by the concavity of the curve, defining the curve to be concave if the difference of the increments ddv (which we would now write as $(d^2v/dx^2) dx^2$) was positive, so that the increments dv themselves increased with increasing v . He defined a point where the increments changed from decreasing to increasing to be a *point of reversed bending* (*punctum flexus contrarii*, what we now call an inflection point), and remarked that at such a point (if it was a point where $dv = 0$ also), the equation had a triple root. What he said is easily translated into the language of today, by looking at the equation $0 = f(x + h) - f(x)$. Obviously, $h = 0$ is a root. At a maximum or minimum, it is a double root. If the point x yields $df = 0$ (that is, $f'(x) = 0$) but is not a maximum or minimum, then $h = 0$ is a triple root ($f''(x) = 0$).

In 1686, Leibniz was the first to use the phrase *osculating circle*. He explained the matter thus:

In the infinitely small parts of a curve it is possible to consider not only the direction or inclination or declination, as has been done up to now, but also the change in direction or curvature (*flexura*), and as the measures of the direction of curves are the simplest lines of geometry having the same direction at the same point, that is, the tangent lines, likewise the measure of curvature is the simplest curve having at the same point not only the same direction but also the same curvature, that is a circle not only tangent to the given curve but, what is more, osculating.³

Leibniz recognized the problem of finding the evolute as that of constructing “not merely an arbitrary tangent to a single curve at an arbitrary point, but a unique common tangent⁴ of infinitely many curves belonging to the same order.” That meant differentiating with respect to the parameter and eliminating it between the equation of the family and the differentiated equation. In short, Leibniz was the first to discuss what is now called the *envelope* of a

³Literally, *kissing*.

⁴The tangent was not necessarily to be a straight line.

family of curves defined by an equation containing a parameter, that is, a curve tangent to every curve in the family that it intersects.

39.2. THE EIGHTEENTH CENTURY: SURFACES

Compared to calculus, differential equations, and analysis in general, differential geometry was not the subject of a large number of papers in the eighteenth century. Nevertheless, there were important advances.

39.2.1. Euler

According to Coolidge (1940, p. 325), Euler's most important contribution to differential geometry came in a 1760 paper on the curvature of surfaces. In that paper he observed that different planes cutting a surface at a point would generally intersect it in curves having different curvatures, but that the two planes for which this curvature was maximal or minimal would be at right angles to each other. For any other plane, making angle α with one of these planes, the radius of curvature would be

$$r = \frac{2fg}{f + g + (g - f) \cos 2\alpha},$$

where f and g are the minimum and maximum radii of curvature at the point. Nowadays, because of an 1813 treatise of Pierre Dupin (1784–1873), this formula is written in terms of the curvature $1/r$ as

$$\frac{1}{r} = \frac{\cos^2 \alpha}{g} + \frac{\sin^2 \alpha}{f},$$

where α is the angle between the given cutting plane and the plane in which the curvature is minimal ($1/g$). The equation obviously implies that in a plane perpendicular to the given plane the curvature would be the same expression with the cosine and sine reversed, or, what is the same, with f and g reversed. Gauss used the formula in Dupin's form, writing it as a formula for the curvature:

$$\frac{1}{T \cos^2 \varphi + V \sin^2 \varphi},$$

where T and V are the maximal and minimal radii of curvature.

Another innovation due to Euler was now-familiar idea of a parameterized surface, in a 1770 paper on surfaces that can be mapped into a plane. The canvas on which an artist paints and the paper on which an engineer or architect draws plans are not only two-dimensional but also *flat*, having curvature zero. Parameters allow the mathematician or engineer to represent information about any curved surface in the form of functions $(t, u) \mapsto (x(t, u), y(t, u), z(t, u))$. Quantities such as curvature and area are then expressed as functions of the parameters (t, u) .

39.2.2. Lagrange

Another study of surfaces, actually a paper in the calculus of variations, was Lagrange's 1762 work on extremal values of integrals.⁵ The connection with differential geometry is in the problem of minimal surfaces and isoperimetric problems, although he began with the brachistochrone problem (finding the curve of most rapid descent for a falling body). Lagrange found a necessary condition for a surface $z = f(x, y)$ to be the minimal surface having a prescribed boundary.

39.3. SPACE CURVES: THE FRENCH GEOMETERS

After these “preliminaries” we finally arrive at the traditional beginning of differential geometry, a 1771 paper of Gaspard Monge on curves in space and his 1780 paper on curved surfaces. Monge elaborated Leibniz' idea for finding the envelope⁶ of a family of lines, considering a family of planes parameterized by their intersections with the z -axis, and obtained the equation of the surface that is the envelope of the family of planes and can be locally mapped into a plane without stretching or shrinking.

39.4. GAUSS: GEODESICS AND DEVELOPABLE SURFACES

With the nineteenth century, differential geometry entered on a period of growth and has continued to reach new heights for two full centuries. The first mathematician to be mentioned is Gauss, who during the 1820s was involved in mapping the region of Hannover in Lower Saxony, where Göttingen is located. This mapping had been ordered by King George IV of England, who was also Elector of Hannover by inheritance from his great grandfather George I. Gauss had been interested in geodesy for many years (Reich, 1977, pp. 29–34) and had written a paper in response to a problem posed by the Danish Academy of Sciences. This paper, which was published in 1825, discussed conformal mapping, that is, mappings that are a pure magnification at each point, so that angles are preserved and the limiting ratio of the actual distance between two points to the map distance between them as one of them approaches the other is the same for approach from any direction.

Involvement with the mapping project inspired Gauss to reflect on the mathematical aspects of developing a curved surface on a flat page and eventually, the more general problem of developing one curved surface on another—that is, mapping the surfaces so that the ratio that the distance from a given point P to a nearby point Q has to the distance between their images P' and Q' tends to 1 as Q tends to P . Gauss apparently planned a full-scale treatise on geodesy but never completed it. Two versions of his major work *Disquisitiones generales circa superficies curvas* (*General Investigations of Curved Surfaces*) were written in the years 1825 and 1827. In the preface to the latter Gauss explained the problem he had set: “to find all representations of a given surface upon another in which the smallest elements remain unchanged.” He admitted that some of what he was doing needed to be made more precise through a more careful statement of hypotheses, but wished to show certain results

⁵*Œuvres de Lagrange*, T. 1, pp. 335–362.

⁶The envelope of a family of surfaces is a surface that is tangent to each of them.

of fundamental importance in the general problem of mapping. He mentioned three ways of defining a surface: first, as the zero set of a function $W(x, y, z)$ of three variables, second as a parameterized mapping $(p, q) \mapsto (x(p, q), y(p, q), z(p, q))$, and third as the graph of a function $z = f(x, y)$. The third case, he pointed out, was merely a specialization of either of the first two.

To determine the extent to which a surface curves, Gauss represented any line in space by a point on a fixed sphere of unit radius: the endpoint of the radius parallel to the line.⁷ This idea, he said, was inspired by the use of the celestial sphere in geometric astronomy. When the line is the normal line through a point of the surface, the result is a mapping from the surface to the unit sphere, so that the sphere and the surface have parallel normal lines at corresponding points. Obviously a plane maps to a single point under this procedure, since all of its normal lines are parallel to one another. Gauss proposed to use the area of the portion of the sphere covered by this map as a measure of curvature of the surface in question. He called this area the *total curvature* of the surface. He then attached a sign to this total curvature by specifying that it was to be positive if the surface was convex in both of two mutually perpendicular directions and negative if it was convex in one direction and concave in the other (like a saddle). Gauss gave an informal discussion of this question in terms of the side of the surface on which an oriented normal line was pointing. When the quality of convexity varied in different parts of a surface, Gauss said, a still more refined definition was necessary, which he found it necessary to omit. Along with the total curvature he defined what we would call its density function and he called the *measure of curvature*, namely the ratio of the area of a local neighborhood on the sphere to the area of the local neighborhood on the surface corresponding to the same parameter values under the two mappings. He denoted this measure of curvature k , a positive or negative sign being attached in accordance with the principles mentioned above. The simplest example is provided by a sphere of radius R , any region of which projects to the similar region on the unit sphere. The ratio of the areas is $k = 1/R^2$, which is therefore the measure of curvature of a sphere at every point.

Thus, in discussing curvature when the surface is given by parameters, Gauss used two mappings from the parameter space (p, q) into three-dimensional space. The first was the mapping onto the surface itself:

$$(p, q) \mapsto (x(p, q), y(p, q), z(p, q)).$$

The second was the mapping

$$(p, q) \mapsto (X(p, q), Y(p, q), Z(p, q))$$

to the unit sphere, which takes (p, q) to the three direction cosines of the normal to the surface at the point $(x(p, q), y(p, q), z(p, q))$.

From these preliminaries, Gauss was able to derive very simply what he himself described as “almost everything that the illustrious Euler was the first to prove about the curvature of curved surfaces.” In particular, he showed that his measure of curvature k was the reciprocal of the product of the two principal radii of curvature that Euler called f and g . He then went

⁷An oriented line is meant here, since there are obviously two opposite radii parallel to the line. Gauss surely knew that the order of the parameters could be used to fix this orientation.

on to consider more general parameterized surfaces. Here he introduced the now-standard quantities E , F , and G , given by

$$\begin{aligned} E &= \left(\frac{\partial x}{\partial p}\right)^2 + \left(\frac{\partial y}{\partial p}\right)^2 + \left(\frac{\partial z}{\partial p}\right)^2, \\ F &= \frac{\partial x}{\partial p} \frac{\partial x}{\partial q} + \frac{\partial y}{\partial p} \frac{\partial y}{\partial q} + \frac{\partial z}{\partial p} \frac{\partial z}{\partial q}, \\ G &= \left(\frac{\partial x}{\partial q}\right)^2 + \left(\frac{\partial y}{\partial q}\right)^2 + \left(\frac{\partial z}{\partial q}\right)^2, \end{aligned}$$

and what is now called the *first fundamental form* for the square of the element of arc length:

$$ds^2 = E dp^2 + 2F dp dq + G dq^2.$$

It is easy to compute that the element of area—the area of an infinitesimal parallelogram whose sides are $(\frac{\partial x}{\partial p} dp, \frac{\partial y}{\partial p} dp, \frac{\partial z}{\partial p} dp)$ and $(\frac{\partial x}{\partial q} dq, \frac{\partial y}{\partial q} dq, \frac{\partial z}{\partial q} dq)$ —is just $\Delta dp dq$, where $\Delta = \sqrt{EG - F^2}$. Gauss denoted the analogous expression for the mapping into the unit sphere $((p, q) \mapsto (X(p, q), Y(p, q), Z(p, q)))$, by

$$D dp^2 + 2D' dp dq + D'' dq^2. \quad (39.1)$$

This quadratic form—or a multiple of it, since definitions vary—is called the *second fundamental form*. As just described, it is produced using the mapping into the unit sphere to generate the element of surface area on that sphere in terms of the parameters. This parameterization can be used to generate an oriented normal line, which must be parallel to the line from the origin to the image point on the unit sphere where the normal is calculated. If that normal points outward from the unit sphere, the curvature of the surface at the corresponding point is positive. If it points inward, that curvature is negative.

The element of area on the unit sphere is $\sqrt{DD'' - (D')^2} dp dq$. Hence, up to the choice of sign, the measure of curvature—what is now called the *Gaussian curvature* and denoted k —is

$$\sqrt{\frac{DD'' - (D')^2}{EG - F^2}}.$$

In a very prescient remark that was later to be developed by Riemann, Gauss noted that “for finding the measure of curvature, there is no need of finite formulæ, which express the coordinates x , y , z as functions of the indeterminates p , q ; but that the general expression for the magnitude of any linear element is sufficient.” The idea is that the geometry of a surface is to be built up from the infinitesimal level using the parameters, not derived from the metric imposed on it by its position in Euclidean space. That is the essential idea of what is now called a *differentiable manifold*.

Summary of the History up to this Point. From the work of Euler, it was clear that a critical point of a function $z = f(x, y)$ (a point (x, y) where the two partial derivatives are zero) is an extremum if the two principal radii of curvature at the point have the same sign

and a saddle point if they have opposite sign. Gauss gave an interpretation to the absolute value of the product of these two curvatures as the limit of the ratio of the area of a small patch of surface near the point to the area of the trace of its unit normal over that patch. Thus geometry and analysis came together very fruitfully in this work. The connection between area and curvature was to have profound consequences.

39.4.1. Further Work by Gauss

It is also clear from Gauss' correspondence (Klein, 1926, p. 16) that Gauss already realized that non-Euclidean geometry was consistent. In fact, the question of consistency did not trouble him; he was more interested in measuring large triangles to see if the sum of their angles could be demonstrably less than two right angles. If so, what we now call hyperbolic geometry might be more convenient for physics than Euclidean geometry.

Gauss considered the possibility of developing one surface on another, that is, mapping it in such a way that lengths are preserved on the infinitesimal level. If the mapping is $(x, y, z) \mapsto (u, v, w)$, then by composition, u , v , and w are all functions of the same parameters that determine x , y , and z , and they generate functions E' , F' , and G' for the second surface that must be equal to E , F , and G at the corresponding points, since that is what is meant by developing one surface on another. But since he had just derived an expression for the measure of curvature that depended only on E , F , G and their partial derivatives, he was able to state the profound result that has come to be called his *theorema egregium* (*outstanding theorem*): *If a curved surface is developed on any other surface, the measure of curvature at each point remains unchanged.* This theorem implies that surfaces that can be developed on a plane, such as a cone or cylinder, must have Gaussian curvature 0 at each point.

With the first fundamental form Gauss was able to derive a pair of differential equations that must be satisfied by geodesic lines, which he called *shortest lines*,⁸ and prove that a geodesic circle—the set of endpoints of geodesics originating at a given point and having a given length—intersects each geodesic at a right angle. This result was the foundation for a generalized theory of polar coordinates on a surface, using p as the distance along a geodesic from a variable point to a pole of reference and q as the angle between that geodesic and a fixed geodesic through the pole. This topic very naturally led to the subject of geodesic triangles, formed by joining three points to one another along geodesics. Since he had shown earlier that the element of surface area was

$$d\sigma = \sqrt{EG - F^2} dp dq,$$

and that this expression was particularly simple when one of the sets of coordinate lines consisted of geodesics (as in the case of a sphere, where the lines of longitude are geodesics), the total curvature of such a triangle was easily found for a geodesic triangle and turned out to be

$$A + B + C - \pi,$$

where A , B , and C are the angles of the triangle, expressed in radians. For a plane triangle this expression is zero. For a spherical triangle it is, not surprisingly, the area of the triangle

⁸According to Klein (1926, Vol. 2, p. 148), the term *geodesic* was first used by Joseph Liouville (1809–1882) in 1850. Klein cites an 1893 history of the term by Paul Stäckel (1862–1919) as source.

divided by the square of the radius of the sphere. In this way, area, curvature, and the sum of the angles of a triangle were shown to be linked on curved surfaces. This result was the earliest theorem on global differential geometry, since it applies to any surface that can be triangulated. In its modern version, it relates curvature to the topological property of the surface as a whole known as the *Euler characteristic* mentioned in the previous chapter. It is called the *Gauss–Bonnet theorem* after Pierre Ossian Bonnet (1819–1892), who introduced the notion of the geodesic curvature of a curve on a surface (that is, the tangential component of the acceleration of a point moving along the curve with unit speed)⁹ and generalized the formula to include this concept.

39.5. THE FRENCH AND BRITISH GEOMETERS

In France, differential geometry was of interest for a number of reasons connected with physics. In particular, it seemed applicable to the problem of heat conduction, the theory of which had been pioneered by such outstanding mathematicians as Jean-Baptiste Joseph Fourier (1768–1830), Siméon-Denis Poisson (1781–1840), and Gabriel Lamé (1795–1870), since isothermal surfaces and curves in a body were a topic of primary interest. It also applied to the theory of elasticity, studied by Lamé and Sophie Germain (1776–1831), among others. Lamé developed a theory of elastic waves that he hoped would explain light propagation in an elastic medium called ether. Sophie Germain noted that the average of the two principal curvatures derived by Euler would be the same for any two mutually perpendicular planes cutting a surface. She therefore recommended this average curvature as the best measure of curvature. Her idea is useful in elasticity theory,¹⁰ but turns out not to be so useful for pure geometry.¹¹ Joseph Liouville (1809–1882) proved that conformal maps of three-dimensional regions are far less varied than those in two dimensions, being necessarily either inversions or similarities or rigid motions. He published this result in the fifth edition of Monge’s book on the applications of analysis to geometry. In contrast, a mapping $(x, y) \mapsto (u, v)$ is conformal if and only if one of the functions $u(x, y) \pm iv(x, y)$ is analytic. As a consequence, there is a rich supply of conformal mappings of the plane.

After Newton, differential geometry languished in Britain until the nineteenth century, when William Rowan Hamilton (1805–1865) published papers on systems of rays, building the foundation for the application of differential geometry to differential equations. Another British mathematician, George Salmon (1819–1904), made the entire subject more accessible with his famous textbooks *Higher Plane Curves* (1852) and *Analytic Geometry of Three Dimensions* (1862).

39.6. GRASSMANN AND RIEMANN: MANIFOLDS

Once the idea of using parameters to describe a surface has been grasped, the development of geometry can proceed algebraically, without reference to what is possible in three-

⁹According to Struik (1933, 20, pp. 163, 165), even this concept was anticipated by Gauss in an unpublished paper of 1825 and followed up on by Ferdinand Minding (1806–1885) in a paper in the *Journal für die reine und angewandte Mathematik* in 1830.

¹⁰In particular, her concept of the average curvature plays a role in the Navier–Stokes equations.

¹¹The average curvature must be zero on a minimal surface, however.

dimensional Euclidean space. This idea was developed in the mid-nineteenth century by a number of German and Italian mathematicians.

39.6.1. Grassmann

One mathematician who took the algebraic point of view in geometry was Hermann Grassmann (1809–1877), a secondary-school teacher, who wrote a philosophically oriented mathematical work published in 1844 under the title *Die lineale Ausdehnungslehre, ein neuer Zweig der Mathematik* (*The Theory of Linear Extensions, a New Branch of Mathematics*). This work, which developed ideas Grassmann had conceived earlier in a work on the ebb and flow of tides, contained much of what is now regarded as multilinear algebra. What we call the coefficients in a linear combination of vectors Grassmann called the numbers by means of which the quantity was derived from the other quantities. He introduced what we now call the tensor product and the wedge product for what he called extensive quantities. He referred to the tensor product simply as the *product* and the wedge product as the *combinatory product*. The tensor product of two extensive quantities $\sum \alpha_r e_r$ and $\sum \beta_s e_s$ was

$$\left[\sum_r \alpha_r e_r, \sum_s \beta_s e_s \right] = \sum_{r,s} \alpha_r \beta_s [e_r, e_s].$$

The combinatory product was obtained by applying to this product the rule that $[e_r, e_s] = -[e_s, e_r]$ (antisymmetrizing). The determinant is a special case of the combinatory product. Grassmann remarked that when the factors are “numerically related” (which we call linearly dependent), the combinatory product would be zero. When the basic units e_r and e_s were entirely distinct, Grassmann called the combinatory product the *outer product* to distinguish it from the *inner product*, which is still called by that name today and amounts to the ordinary dot product when applied to vectors in physics. Grassmann remarked that parentheses have no effect on the outer product—in our terms, it is an associative operation.¹²

Working with these concepts, Grassmann defined the *numerical value* of an extended quantity as the positive square root of its inner square, exactly what we now call the absolute value of a vector in n -dimensional space. He proved that “the quantities of an orthogonal system are not related numerically,” that is, an orthogonal set of nonzero vectors is linearly independent.

39.6.2. Riemann

Historians of mathematics seem to agree that, because of its philosophical tone and unusual nomenclature, *Ausdehnungslehre* did not attract a great deal of notice until Grassmann revised it and published a more systematic exposition in 1862. If that verdict is correct, there is a small coincidence in Riemann’s use of the term “extended,” which appears to mimic Grassmann’s use of the word, and in his focus on a general number of dimensions in his inaugural lecture at the University of Göttingen.

¹²To avoid confusing the reader who knows that the cross product is not an associative product, we note that the outer product applies only when each of the factors is orthogonal to the others. In three dimensional space the cross product of three such vectors, however they are grouped, is always zero. The wedge product, however, is associative.

Riemann's most authoritative biographer Laugwitz (1999, p. 223) says that Grassmann's work would have been of little use to Riemann, since for him linear algebra was a trivial subject.¹³ The inaugural lecture was read in 1854, with the aged Gauss in the audience.¹⁴ Although Riemann's lecture "Über die Hypothesen die der Geometrie zu Grunde liegen" ("On the hypotheses that form the basis of geometry") occupies only 14 printed pages and contains almost no mathematical symbolism—it was aimed at a largely nonmathematical audience—it set forth ideas that had profound consequences for the future of both mathematics and physics. As Hermann Weyl said:

The same step was taken here that was taken by Faraday and Maxwell in physics, the theory of electricity in particular, . . . by passing from the theory of action at a distance to the theory of local action: the principle of understanding the world from its behavior on the infinitesimal level. [Narasimhan, 1990, p. 740]

In the first section, Riemann began by developing the concept of an n -fold extended quantity, asking the indulgence of his audience for delving into philosophy, where he had limited experience. He cited only some philosophical work of Gauss and of Johann Friedrich Herbart, who was mentioned in the previous chapter, Riemann began with the concept of quantity in general, which arises when some general concept can be defined (measured or counted) in different ways. Then, according as there is or is not a continuous transformation from one of the ways into another, the various determinations of it form a continuous or discrete manifold. He noted that discrete manifolds (sets of things that can be counted, as we would say) are very common in everyday life, but continuous manifolds are rare, the spatial location of objects of sense and colors being almost the only examples.

The main part of the lecture was the second part, in which Riemann investigated the kinds of metric relations that could exist in a manifold if the length of a curve was to be independent of its position. Assuming that the point was located by a set of n coordinates x_1, \dots, x_n (almost the only mathematical symbols that appear in the paper), he considered the kinds of properties needed to define an infinitesimal element of arc length ds along a curve. The simplest function that met this requirement was

$$ds = \sqrt{\sum a_{ij}(x_1, \dots, x_n) dx_i dx_j},$$

where the coefficients a_{ij} were continuous functions of position and the expression under the square root is always nonnegative. The next simplest case, which he chose not to develop, occurred when the Maclaurin series began with fourth-degree terms. As Riemann said,

The investigation of this more general type, to be sure, would not require any essentially different principles, but it would be rather time-consuming and cast relatively little new light on the theory of space; and moreover the results could not be expressed geometrically.

¹³One can't help wondering about the *multilinear* algebra that Grassmann was developing. The recognition of this theory as an essential part of geometry is explicit in Felix Klein's 1908 work on elementary geometry from a higher viewpoint, but Riemann apparently did not make the connection.

¹⁴At the time of the lecture, Gauss had less than a year of life remaining. Yet his mind was still active, and he was very favorably impressed by Riemann's performance.

For the case in which coordinates could be chosen so that $a_{ii} = 1$ and $a_{ij} = 0$ when $i \neq j$, Riemann called the manifold *flat*.

Having listed the kinds of properties space was assumed to have, Riemann asked to what extent these properties could be verified by experiment, especially in the case of continuous manifolds. What he said at this point has become famous. He made a distinction between the infinite and the unbounded, pointing out that while space is always assumed to be unbounded (that is, to have no *border*), it might very well not be infinite. Then, as he said, assuming that solid bodies exist independently of their position, it followed that the curvature of space would have to be constant, and all astronomical observation confirmed that it could only be zero. But, if the volume occupied by a body varied as the body moved, no conclusion about the infinitesimal nature of space could be drawn from observations of the metric relations that hold on the finite level. "It is therefore quite conceivable that the metric relations of space are not in agreement with the assumptions of geometry, and one must indeed assume this if phenomena can be explained more simply thereby." Such reasoning plays a role in the theory of relativity, where the rigid body of classical mechanics does not exist.

Riemann evidently intended to follow up on these ideas, but his mind produced ideas much faster than his frail body would allow him to develop them. He died before his 40th birthday with this project one of many left unfinished. He did, however, send an essay to the Paris Academy in response to a prize question proposed (and later withdrawn): *Determine the thermal state of a body necessary in order for a system of initially isothermal lines to remain isothermal at all times, so that its thermal state can be expressed as a function of time and two other variables*. Riemann's essay was not awarded the prize because its results were not developed with sufficient rigor. It was not published during his lifetime.¹⁵

39.7. DIFFERENTIAL GEOMETRY AND PHYSICS

The work of Grassmann and Riemann was to have a powerful impact on the development of both geometry and physics. One has only to read Einstein's accounts of the development of general relativity to understand the extent to which he was imbued with Riemann's outlook. The idea of geometrizing physics seems an attractive one. The Aristotelian idea of force, which had continued to serve through Newton's time, began to be replaced by subtler ideas developed by the Continental mathematical physicists of the nineteenth century, with the introduction of such principles as conservation of energy and least action. In his 1736 treatise on mechanics, Euler had shown that a particle constrained to move along a surface by forces normal to the surface, but on which no forces tangential to the surface act, would move along a shortest curve on the surface. And when he discovered the variational principles that enabled him to solve the isoperimetric problem, he applied them to the theory of elasticity and vibrating membranes. As he said,

Since the material of the universe is the most perfect and proceeds from a supremely wise Creator, nothing at all is found in the world that does not illustrate some maximal or minimal principle. For that reason, there is absolutely no doubt that everything in the universe, being the result of an ultimate purpose, is amenable to determination with equal success from these efficient causes using the method of maxima and minima. [Euler, 1744, p. 245]

¹⁵Klein (1926, Vol. 2, p. 165) notes that very valuable results were often submitted for prizes at that time, since professors were so poorly paid.

Riemann was searching for connections among light, electricity, magnetism, and gravitation at this time.¹⁶ In 1846, Gauss' collaborator Wilhelm Weber (1804–1891) had incorporated the velocity of light in a formula for the force between two moving charged particles. According to Hermann Weyl (Narasimhan, 1990, p. 741), Riemann did not make any connection between that search and the content of his inaugural lecture. Laugwitz (1999, p. 222), however, cites letters from Riemann to his brother which show that he did make precisely that connection. Whatever the case, four years later Riemann sent a paper¹⁷ to the Royal Society in Göttingen in which he made the following remarkable statement:

I venture to communicate to the Royal Society a remark that brings the theory of electricity and magnetism into a close connection with the theory of light and heat radiation. I have found that the electrodynamic effects of galvanic currents can be understood by assuming that the effect of one quantity of electricity on others is not instantaneous but propagates to them with a velocity that is constant (equal to that of light within observational error).

39.8. THE ITALIAN GEOMETERS

The political unification of Italy in the mid-nineteenth century was accompanied by a surge of mathematical activity even greater than the sixteenth-century work in algebra. Gauss had analyzed a general surface by using two parameters and introducing six functions: the coefficients of the first and second fundamental forms. The question naturally arises whether a surface can be synthesized from any six functions regarded as the coefficients of these forms. Do they determine the surface, up to the usual Euclidean motions of translation, rotation, and reflection that can be used to move a set of axes to a prescribed position and orientation? Such a theorem does hold for curves, as was established by two French mathematicians, Jean Frenet (1816–1900) and Joseph Serret (1818–1885), who gave a set of equations—the Frenet–Serret¹⁸ equations—determining the curvature and torsion¹⁹ of a curve in three-dimensional spaces. A curve can be reconstructed from its curvature and torsion up to translation, rotation, and reflection. A natural related question is: Which sets of six functions, regarded as the components of the two fundamental forms, can be used to construct a surface? After all, one needs generally only three functions of two parameters to determine a surface, so that the six given by Gauss cannot be independent of one another.

In an 1856 paper, Gaspare Mainardi (1800–1879) provided consistency conditions in the form of four differential equations, now known as the Mainardi–Codazzi equations,²⁰ which must be satisfied by the six functions E , F , G , D , D' , and D'' if they are to be the components of the first and second fundamental forms introduced by Gauss. Mainardi had

¹⁶His lecture was given nearly a decade before Maxwell discovered his famous equations connecting the speed of light with the propagation of electromagnetic waves.

¹⁷This paper was later withdrawn, but was published after his death (Narasimhan, 1990, pp. 288–293).

¹⁸Frenet gave six equations for the direction cosines of the tangent and principal normal to the curve and its radius of curvature. Serret gave the full set of nine now called by this name, which are more symmetric but contain no more information than the six of Frenet.

¹⁹The torsion of a curve measures its tendency to move out of the plane of its tangential and principal normal vectors.

²⁰The Latvian mathematician Karl Mikhailovich Peterson (1828–1881) published an equivalent set of equations in Moscow in 1853, but they went unnoticed for a full century.

learned of Gauss' work through a French translation, which had appeared in 1852. These same equations were discovered by Delfino Codazzi (1824–1875) two years later, using an entirely different approach, and helped him to win a prize from the Paris Academy of Sciences. Codazzi published these equations only in 1883.

When Riemann's lecture was published in 1867, the year after his death, it became the point of departure for a great deal of research in Italy.²¹ One who worked to develop these ideas was Riemann's friend Enrico Betti, who tried to get Riemann a chair of mathematics in Palermo. These ideas led Betti to the notion of the connectivity of a surface. On the simplest surfaces, such as a sphere, every closed curve is the boundary of a region. On a torus, however, the circles of latitude and longitude are not boundaries. These ideas belong properly to topology, and were discussed in the preceding chapter. In his fundamental work on this subject, Henri Poincaré named the maximum number of independent nonboundary cycles in a surface the *Betti number* of the surface, a concept that is now generalized to n dimensions. The n th Betti number is the rank of the n th homology group.

Another Italian mathematician who extended Riemann's ideas was Eugenio Beltrami (1835–1900), whose 1868 paper on spaces of constant curvature contained a model of a three-dimensional space of constant negative curvature. Beltrami had previously given the model of a pseudosphere to represent the hyperbolic plane, which will be discussed in the next chapter. It was not obvious before his work that three-dimensional hyperbolic geometry and a three-dimensional manifold of constant negative curvature were basically the same thing. Beltrami also worked out the appropriate n -dimensional analogue of the Laplacian $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}$, which plays a fundamental role in mathematical physics. By working with an integral considered earlier by Jacobi (see Klein, 1926, Vol. 2, p. 190), Beltrami arrived at the operator

$$\Delta u = \frac{1}{\sqrt{a}} \sum_{i=1}^n \frac{\partial}{\partial x^i} \left(\sqrt{a} \sum_{j=1}^n a_{ij} \frac{\partial u}{\partial x^j} \right),$$

where, with the notation slightly modernized, the Riemannian metric is given by the usual $ds^2 = \sum_{i,j=1}^n a_{ij} dx^i dx^j$, and a denotes the determinant $\det(a_{ij})$. The generalized operator is now referred to as the Laplace–Beltrami operator on a Riemannian manifold.

39.8.1. Ricci's Absolute Differential Calculus

The algebra of Grassmann and its connection with Riemann's general metric on an n -dimensional manifold was not fully codified until 1901, in "Méthodes de calcul différentiel absolu et leurs applications" ("Methods of absolute differential calculus and their applications"), published in *Mathematische Annalen* in 1901, written by Gregorio Ricci-Curbastro (1853–1925) and Tullio Levi-Civita (1873–1941). This article contained the critical ideas of tensor analysis as it is now taught. The absoluteness of the calculus consisted in the great generality of the transformations that it permitted, showing how differential forms changed when coordinates were changed. Although Ricci-Curbastro competed in a prize

²¹Riemann went to Italy for his health and died of tuberculosis in Selasca. He was in close contact with Italian mathematicians and even published a paper in Italian.

contest sponsored that year by the Accademia dei Lincei, he was not successful. Some of the judges regarded his absolute differential calculus as superfluous to the end it was designed for.²²

The following year, Luigi Bianchi (1873–1928) published “Sui simboli a quattro indice e sulla curvatura di Riemann” (“On quadruply-indexed symbols and Riemannian curvature”), in which he gave the relations among the covariant derivatives of the Riemann curvature tensor, which he derived by a direct method for manifolds of constant curvature, not following the route of Ricci-Curbastro and Levi-Civita. The Bianchi identity was later to play a crucial role in general relativity, assuring local conservation of energy when Einstein’s gravitational equation is assumed.

PROBLEMS AND QUESTIONS

Mathematical Problems

- 39.1. Find the radius of curvature of the parabola $y = x^2$ at the point $(1, 1)$. (The center of curvature lies along the normal line at that point, which has equation $x + 2y = 3$.)
- 39.2. Find the first fundamental form of the hyperbolic paraboloid $z = (x^2 - y^2)/a$ at each point using x and y as parameters.
- 39.3. Find the Gaussian curvature of the hyperbolic paraboloid described in the previous problem.

Historical Questions

- 39.4. What motives did Huygens and Newton have for studying involutes, evolutes, and curvature?
- 39.5. What significant increase in the algebraization of geometry was promoted by Euler?
- 39.6. In what ways did Riemann anticipate much of modern physics?

Questions for Reflection

- 39.7. We saw earlier (Chapter 27) that Omar Khayyam criticized ibn al-Haytham for discussing a line that moves while remaining perpendicular to another line. Yet Newton used exactly the same language when discussing the center of curvature, and no one seems to have objected. What difference in mathematical cultures does this contrast in points of view signify?
- 39.8. Riemann raised the issue whether space was infinite or finite. In 1976, the mathematician John Milnor (b. 1931), in an address to the Northeastern Section of the Mathematical Association of America, asked whether space is simply connected. One way of investigating this question would be to look for doubly periodic star patterns, such as one would see if the celestial sphere were actually a torus. In the

²²The same sort of criticism was leveled by Weierstrass against the work of Hamilton in quaternions.

discussion that followed, a member of the audience asked what reason we have for thinking that space is even orientable. Might it not resemble a projective plane? How could one investigate this question empirically?

- 39.9.** How much of modern differential geometry could be cast in the language of Euclidean geometry? How would you describe, for example, the hyperbolic paraboloid whose equation is $az = x^2 - y^2$?

Non-Euclidean Geometry

There are two opposite points of view that one might adopt when assessing the value of Euclid's axiomatic approach to geometry. On the one hand, one could argue that it was precisely the attempt to spell out all assumptions explicitly that led to an explicit statement of the parallel postulate and hence subjected it to scrutiny. Thus, Euclidean geometry really generated non-Euclidean. That fact is underscored by the total absence of any speculation along those lines by mathematical cultures not descended from Euclid. On the other hand, it was the Hindu–Arabic algebra that immensely increased the power of geometry through analytic, projective, algebraic, and differential geometry. This algebraic approach laid out myriad examples of non-Euclidean geometries. All one had to do was look at them to see the possibility of denying the parallel postulate. From that point of view, Euclid's geometry is merely one specimen among many, all of roughly equal value for science, and “non-Euclidean” geometry is an unnecessary name for all the other surfaces and manifolds that don't happen to be Euclidean.

The hold that Euclid had over the intellectual imagination of the West was vast in its extent. For centuries, the axiomatic approach to all kinds of knowledge was regarded as an ideal in every area of intellectual endeavor. The philosopher Baruch Spinoza (1632–1677), for example, wrote a book entitled *Ethica ordine geometrico demonstrata*.¹ From the time of Descartes on, mathematicians had found the algebra inherited from the Hindus and Muslims and developed into a powerful symbolic method to be far superior, and they were the first to turn to other methods and let Euclid fall into neglect. Yet even after differential geometry was well established and a variety of exotic surfaces became amenable to study, a few mathematicians were still treading the old Euclidean paths and trying to prove the parallel postulate. In the nineteenth century, several of these people became the pioneers of non-Euclidean geometry, developed axiomatically, just like Euclidean. However, they took advantage of the metric point of view and developed the trigonometry of their non-Euclidean planes. It was this infusion of algebra into their reasoning that eventually brought about the acceptance of their work. At the same time, despite the repairs made to the Euclidean system by Hilbert, the presence of other geometries with equal claim to validity at long last settled not only the question of the provability of that postulate, but made the whole edifice of Euclidean methodology irrelevant to mathematical progress. The work of Euclid,

¹*Ethics Demonstrated in Geometric Order*. It was published a few months after Spinoza's death.

Archimedes, and Apollonius was no less remarkable for this shift in point of view. It remains an imposing intellectual achievement, just as the great stone monuments that continue to fascinate the imagination of modern people remain as a testimony to the genius of the builders. But no architect nowadays looks to Egyptian building methods for ideas or holds them up as a model to be imitated. Euclid and his contemporaries have a permanent place in our culture. A liberal education that attempts to acquaint students with the greatest human achievements should not neglect them. But they also should not be presented as the sum total of mathematical achievement. Non-Euclidean geometry has traditionally meant geometry with the parallel postulate replaced by some other postulate. It really should be taken to mean geometry that is simply independent of, and draws on other sources in addition to, Euclid, that is to say, 99% of what is now called geometry.

The centuries of effort by Hellenistic and Islamic mathematicians to establish the parallel postulate as a fact of nature began to be repeated in early modern Europe, as mathematicians tried to replace the postulate with some other assumption that seemed more obvious. Then, around the year 1800, a change in attitude took place, as a few mathematicians began to explore non-Euclidean geometries as if they might have some meaning after all. Within a few decades the full light of day dawned on this topic, and by the late nineteenth century, models of the non-Euclidean geometries inside Euclidean and projective geometry removed all doubt as to their consistency. This history exhibits a sort of parallelism (no pun intended) with the history of the classical construction problems and with the problem of solving higher-degree equations in radicals, all of which were shown in the early nineteenth century to be impossible tasks. In all three cases, group theory eventually played a role in understanding the issues.

40.1. SACCHERI

The Jesuit priest Giovanni Saccheri (1667–1733), a professor of mathematics at the University of Pavia, published in the last year of his life the treatise *Euclides ab omni naevo vindicatus (Euclid Acquitted of Every Blemish)*, a good example of the creativity a very intelligent person will exhibit when trying to retain a strongly held belief. Some of his treatise duplicated what had already been done by the Islamic mathematicians, including the study of Thabit quadrilaterals, that is, quadrilaterals having a pair of equal opposite sides and equal base angles or having having three right angles. Saccheri deduced with strict rigor all the basic properties of Thabit quadrilaterals with right angles at the base (see Chapter 27).² He realized that the fundamental question involved the summit angles of these quadrilaterals—Saccheri quadrilaterals, as they are now called. Since these angles were equal, the only question was whether they were obtuse, right, or acute angles. He showed in Propositions 5 and 6 that if one such quadrilateral had obtuse summit angles, then all of them did likewise, and that if one had right angles, then all of them did likewise. It followed by elimination and without further proof (Proposition 7, which Saccheri proved anyway) that if one of them had acute angles, then all of them did likewise. Not being concerned to

²It is unlikely that Saccheri knew of the earlier work by Thabit ibn-Qurra and others. Although Arabic manuscripts stimulated a revival of mathematics in Europe, not all of them became known immediately. Some of those who did were neglected by historians of the subject. Coolidge (1940) gave the history of the parallel postulate, jumping directly from Proclus and Ptolemy to Saccheri, never mentioning any of the Muslim mathematicians.

eliminate the possibility of the right angle, which he believed was the true one, he worked to eliminate the other two hypotheses.

He showed that the postulate as Euclid stated it is true under the hypothesis of the obtuse angle. That is, two lines cut by a transversal in such a way that the interior angles on one side are less than two right angles will meet on that side of the transversal. As we know, that is because they will meet on *both* sides of the transversal, assuming it makes sense to talk of opposite sides. Saccheri remarked that the intersection must occur at a finite distance. Saccheri would soon be reasoning about points at infinity as if something were known about them, even though he had no careful definition of them.

It is true, as many have pointed out, that his proof of this fact uses the exterior angle theorem (Proposition 16 of Book 1 of Euclid) and hence assumes that lines are infinite.³ But Euclid himself, as later edited, states explicitly that two lines cannot enclose an area, so that Saccheri can hardly be faulted for dealing with only one Euclidean postulate at a time. Since the parallel postulate implies that the summit and base of a Saccheri quadrilateral must meet on *both* sides of the quadrilateral under the hypothesis of the obtuse angle, even a severe critic should be inclined to give Saccheri a passing grade when he rejects this hypothesis.

Having disposed of the hypothesis of the obtuse angle, Saccheri then joined battle (his phrase) with the hypothesis of the acute angle. Here again, he proved some basic facts about what we now call hyperbolic geometry. Given any quadrilateral having right angles at the base and acute angles at the summit, it follows from continuity considerations that the length of a perpendicular dropped from the summit to the base must reach a minimum at some point, and at that point it must also be perpendicular to the summit. Saccheri analyzed this situation in detail, describing in the process some of the phenomena that must occur in what is now called hyperbolic geometry. In terms of Fig. 40.1a,⁴ he considered all the lines like AF through the point A such that angle BAF is acute. He wished to show that they all intersected the line BE .

Saccheri proved that there must be at least one angle θ_0 for which the line AL making that angle neither intersects BE nor has a common perpendicular with it. This line, as Saccheri showed in Proposition 23, must approach BE asymptotically as we would say. At that point he made the small slip that had been warned against even in ancient times, assuming that “approaching” implies “meeting.” His intuition for hyperbolic geometry was very good, as he imagined a line UV perpendicular to BE moving away from AB to positions such as HI and the lines AV , AI , and so on from A perpendicular to the moving line rotating clockwise about A to make angles that decreased to θ_0 . He then—too hastily, as we now know—drew the conclusion that θ_0 would have the properties of *both* of the sets of angles that it separated, that is, the line making this angle would intersect BE and would also have a common perpendicular with it. In fact, it has neither property. But Saccheri was determined to have both. As he described the situation, the hypothesis of the acute angle implied the

³Actually, the use of that proposition is confined to elaborations by the modern reader. The proof stated by Saccheri uses only the fact that lines are *unbounded*, that is, can be extended to any length. It is not necessary to require that the extension never overlap the portion already present.

⁴Since the flat page is not measurably non-Euclidean, and wouldn't be even if spread out to cover the entire solar system, the kinds of lines that occur in hyperbolic geometry cannot be drawn accurately on paper. Our convention is the usual one: When asymptotic properties are not involved, draw the lines straight. When asymptotic properties need to be shown, draw them as hyperbolas. Actually, if the radius of curvature of the plane were comparable to the width of the page, two lines with a common perpendicular would diverge from each other like the graphs of $\cosh x$ and $-\cosh x$, very rapidly indeed.

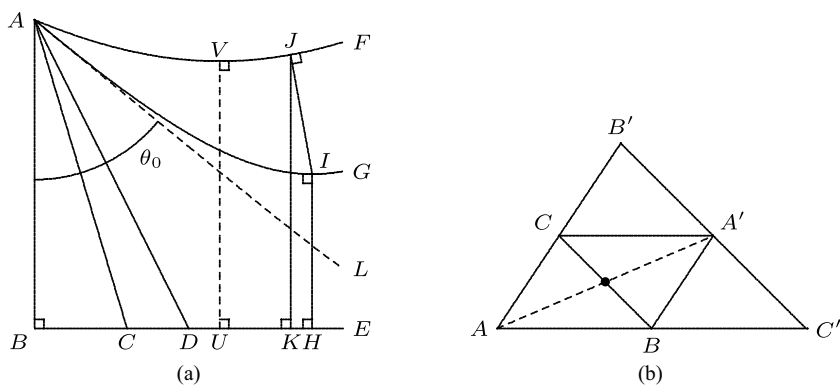


Figure 40.1. (a) Lines like AC and AD through A that intersect BE and those like AF and AG that share a common perpendicular with BE are separated by a line (AL) that is asymptotic to BE . (b) The angle defect of $\triangle AB'C'$ is more than twice the defect of $\triangle ABC$.

existence of two straight lines that have a common perpendicular *at the same point*. In other words, there could be two distinct lines perpendicular to the same line at a point, which is indeed a contradiction. Unfortunately, the point involved was not a point of the plane, but is infinitely distant, as Saccheri himself realized. But he apparently believed that points and lines at infinity must obey the same axioms as those in the finite plane.

Once again, as in the case of Ptolemy, Thabit ibn-Qurra, and ibn al-Haytham, Saccheri had glimpsed a new, non-Euclidean kind of geometry, but resorted to procrustean methods to make it fit his Euclidean intuition.

40.2. LAMBERT AND LEGENDRE

The writings of the Swiss mathematician Johann Heinrich Lambert (1728–1777) seem modern in many ways. For example, he proved that π is irrational (specifically, that $\tan x$ and x cannot both be rational numbers), studied the problem of constructions with straightedge and a fixed compass, and introduced the hyperbolic functions and their identities as they are known today, including the notation $\sinh(x)$ and $\cosh(x)$. He wrote, but did not publish, a treatise on parallel lines, in which he pointed out that the hypothesis of the obtuse angle holds for great circles on a sphere and that the area of a spherical triangle is the excess of its angle sum over π times the square of the radius. He concluded that in a sphere of imaginary radius ir , whose area would be negative, the area of a triangle might be proportional to the excess of π over the angle sum. What a sphere of imaginary radius looks like took some time to discern, a full century, to be exact.

The hyperbolic functions that he studied turned out to be the key to trigonometry in this imaginary world. Just as on the sphere there is a natural unit of length (the radius of the sphere, for example), the same would be true, as Lambert realized, on his imaginary sphere. Such a unit could be selected in a number of ways. The angle of parallelism θ_0 mentioned above, for example, decreases steadily as the length AB increases. Every length is associated with an acute angle, and a natural unit of length might be the one whose angle of parallelism is half of a right angle. Or, it might be the length of the side of an equilateral

triangle having a specified angle. In any case, Lambert at least recognized that he had not proved the parallel postulate. As he said, it was always possible to develop a proof of the postulate to the point that only some small, seemingly obvious point remained unproved, but that last point nearly always concealed an assumption equivalent to what was being proved.

Some of Lambert's reasoning was recast in more precise form by Legendre, who wrote a textbook of geometry used in many places during the nineteenth century, including (in English translation) the United States. Legendre, like Lambert and Saccheri, refuted the possibility that the angle sum of a triangle could be more than two right angles and attempted to show that it could not be less. Since the defect of a triangle—the difference between two right angles and its angle sum—is additive, in the sense that if a triangle is cut into two smaller triangles, the defect of the larger triangle is the sum of the defects of the two smaller ones, he saw correctly that if one could repeatedly double a triangle, eventually the angle sum would have to become negative, which was surely impossible. Unfortunately, the possibility of repeated doubling that he had in mind was just one of those small points mentioned by Lambert that turn out to be equivalent to the parallel postulate. In fact, it is rather easy to see that such is the case, since (Fig. 40.1b) asserting the possibility of drawing a line $B'C'$ through a point A' inside the angle CAB that intersects the extensions of both AB and AC is another way of saying that the line AB cannot be parallel to *every* line through A' that intersects the extension of AC . (If it were, the limiting position of a line through A' intersecting the extension of AC , as the point of intersection tends to infinity in the direction of C from A , would be a line parallel to both AB and AC , and there would thus be two lines through A parallel to this limiting line.)

40.3. GAUSS

The parallel postulate was beginning to be understood by the end of the eighteenth century. Gauss, who read Lambert's work on parallels (which had been published posthumously), began to explore this subject as a teenager, although he kept his thoughts to himself except for letters to colleagues and never published anything on the subject. His work in this area was published in Vol. 8 of the later edition of his collected works. It was summarized by Klein (1926, pp. 58–59). In 1799, Gauss wrote to Farkas Bolyai (1775–1856), his classmate from Göttingen, that he could prove the parallel postulate provided that triangles of arbitrarily large area were admitted. Such a confident statement can only mean that he had developed the metric theory of hyperbolic geometry to a considerable extent. Five years later he wrote again to explain the error in a proof of the parallel postulate proposed by Bolyai. Gauss, like Lambert, realized that a non-Euclidean space would have a natural unit of length, and mentioned this fact in a letter of 1816 to his student Christian Ludwig Gerling (1788–1864), proposing as unit the side of an equilateral triangle whose angles were $59^\circ 59' 59.99999 \dots$.⁵ To Gauss' surprise, in 1818 he received from Gerling a paper written by Ferdinand Karl Schweikart (1780–1859), a lawyer then in Marburg, who had developed what he called *astral geometry*. It was actually hyperbolic geometry, and Schweikart had gone far into it, since he knew that there was an upper bound to the area of a triangle in

⁵In comparison with the radius of curvature of space, this would be an extremely small unit of length. If space is curved negatively at all, however, its radius of curvature is so enormous that this unit would be very large.

this geometry, that its metric properties depended on an undetermined constant C (the distance, measured in units equal to the radius of curvature, at which the angle of parallelism is half of a right angle), and that it contained a natural unit of length, which he described picturesquely by saying that if that length were the radius of the earth, then the line joining two stars would be tangent to the earth. Gauss wrote back to correct some minor points of bad drafting on Schweikart's part (for example, Schweikart neglected to say that the stars were assumed infinitely distant), but generally praising the work. In fact, he communicated his formula for the limiting area of a triangle⁶:

$$\frac{\pi C^2}{(\ln(1 + \sqrt{2}))^2}.$$

By coincidence, Schweikart's nephew Franz Adolph Taurinus (1794–1874), also a lawyer, who surely must have known of his uncle's work in non-Euclidean geometry, sent Gauss his attempt at a proof of the parallel postulate in 1824. Gauss explained the true situation to Taurinus under strict orders to keep the matter secret. The following year, Taurinus published a treatise *Geometria prima elementa* (*First Elements of Geometry*) in which he accepted the possibility of other geometries. Gauss wrote to the astronomer–mathematician Friedrich Wilhelm Bessel (1784–1846) in 1829 that he had been thinking about the foundations of geometry off and on for nearly 40 years (in other words, from the age of 12 on), saying that his investigations were “very extensive,” but probably wouldn't be published, since he feared the controversy that would result. Some time during the mid-1820s, the time when he was writing and publishing his fundamental work on differential geometry, Gauss wrote a note—which, typically, he never published—in which he mentioned that revolving a tractrix about its asymptote produced a surface that is the opposite of a sphere. This surface turns out to be a perfect local model of the non-Euclidean geometry in which the angle sum of a triangle is less than two right angles. It is now called a pseudosphere. This same surface was discussed a decade later by Ferdinand Minding, who pointed out that some pairs of points on this surface can be joined by more than one minimal path, just like antipodal points on a sphere.

40.4. THE FIRST TREATISES

By 1820, the consistency of non-Euclidean geometry was beginning to become plain. As more and more mathematicians worked over the problem and came to the same conclusion, from which others gained insight little by little, all that remained was a slight push to tip the balance from attempts to prove the parallel postulate to the acceptance of alternative hypotheses. The fact that this extra step was taken by several people nearly simultaneously can be expressed poetically, as it was by Felix Klein (1926, p. 57), who referred to “one of the remarkable laws of human history, namely that the times themselves seem to hold the great thoughts and problems and offer them to heads gifted with genius when they are ripe.”

⁶The coefficient of π in this expression represents a unit area and is numerically equal to the square of the radius of curvature of the hyperbolic plane. However, there are no squares in hyperbolic geometry. An equiangular rhombus has acute angles at all four corners.

But we need not be quite so lyrical about a phenomenon that is entirely to be expected: When many intelligent people who have received similar educations work on a problem, it is not surprising when more than one of them makes the same discovery.

The credit for first putting forward hyperbolic geometry for serious consideration must belong to Schweikart, since Gauss was too reticent to do so. However, credit for the first full development of it, including its trigonometry, is due to the Russian mathematician Nikolai Ivanovich Lobachevskii (1792–1856) and the Hungarian János Bolyai (1802–1860), son of Farkas Bolyai. Their approaches to the subject are very similar. Both developed the geometry of the hyperbolic plane and then extended it to three-dimensional space. In three-dimensional space they considered the entire set of directed lines parallel to a given directed line in a given direction. Then they showed that a surface (now called a *horosphere*) that cuts all of these lines at right angles has all the properties of a Euclidean plane. By studying sections of this surface they were able to deduce the trigonometry of their new geometry. The triangle formulas fully justify Lambert's assertion that this kind of geometry is that of a sphere of imaginary radius. Here, for example, is the Pythagorean theorem for a right triangle of sides a, b, c in spherical and hyperbolic geometry, derived by both Lobachevskii and Bolyai, but not in the notation of hyperbolic functions. Since $\cos(ix) = \cosh(x)$ the hyperbolic formula can be obtained from the spherical formula by replacing the radius r with ir , just as Lambert stated.

Spherical geometry	Hyperbolic geometry
$\cos\left(\frac{a}{r}\right) \cos\left(\frac{b}{r}\right) = \cos\left(\frac{c}{r}\right)$	$\cosh\left(\frac{a}{r}\right) \cosh\left(\frac{b}{r}\right) = \cosh\left(\frac{c}{r}\right)$

40.5. LOBACHEVSKII'S GEOMETRY

Lobachevskii connected the parts of a hyperbolic triangle through his formula for the angle of parallelism, which is the angle θ_0 referred to above, as a function of the length AB . He gave this formula as

$$\tan\left(\frac{1}{2}F(\alpha)\right) = e^\alpha,$$

where α denotes the length AB and $F(\alpha)$ the angle θ_0 . Here e could be any positive number, since the radius of curvature of the hyperbolic plane could not be determined. However, Lobachevskii found it convenient to take this constant to be $e = 2.71828\dots$. In effect, he took the radius of curvature of the plane as the unit of length. Lobachevskii gave the Pythagorean theorem, for example, as

$$\sin F(a) \sin F(b) = \sin F(c).$$

Of the two nearly simultaneous creators of hyperbolic geometry and trigonometry, Lobachevskii was the first to publish, unfortunately in a journal of limited circulation. He was a professor at the provincial University of Kazan' in Russia and published his work in 1826 in the proceedings of the Kazan' Physico-Mathematical Society. He reiterated this idea over the next ten years or so, developing its implications. Like Gauss, he drew the conclusion that only observation could determine if actual space was Euclidean or not. As

it happened, astronomers were just beginning to attempt measurements on the interstellar scale. By measuring the angles formed by the lines of sight from the earth to a given fixed star at intervals of six months, one could get the base angles of a gigantic triangle and thereby (since the angle sum could not be larger than two right angles, as everyone agreed) place an upper bound on the size of the parallax of the star (the angle subtended by the earth's orbit from that star). Many encyclopedias claim that the first measurement of stellar parallax was carried out in Königsberg by Gauss' correspondent Bessel in 1838 and that he determined the parallax of 61 Cygni to be 0.3 seconds. Russian historians credit another Friedrich Wilhelm, namely Friedrich Wilhelm Struve (1793–1864), who emigrated to Russia and is known there as Vasilii Yakovlevich Struve. He founded the Pulkovo Observatory in 1839. Struve determined the parallax of the star Vega in 1837. Attempts to determine stellar parallax must have been made earlier, since Lobachevskii cited such measurements in an 1829 work and claimed that the measured parallax was less than $0.000372''$, which is much smaller than any observational error.⁷ As he said (see his collected works, Vol. 1, p. 207, quoted by S. N. Kiro, 1967, Vol. 2, p. 159):

At the very least, astronomical observations prove that all the lines amenable to our measurements, even the distances between celestial bodies, are so small in comparison with the length taken as a unit in our theory that the equations of (Euclidean) plane trigonometry, which have been used up to now must be true without any sensible error.

Thus, the acceptance of the consistency of hyperbolic geometry was accompanied by the rejection of any practical application of it in astronomy or physics. That situation was to change in the early twentieth century, with the advent of relativity.

Lobachevskii was unaware of the work of Gauss, since Gauss kept it to himself and urged others to do likewise. Had Gauss been more talkative, Lobachevskii would easily have found out about his work, since his teacher Johann Martin Christian Bartels (1769–1836) had been many years earlier a teacher of the 8-year-old Gauss and had remained a friend of Gauss. As it was, however, although Lobachevskii continued to perfect his “imaginary geometry,” as he called it, and wrote other mathematical papers, he made his career in administration, as rector of the University of Kazan'. He at least won some recognition for his achievement during his lifetime, and his writings were translated into French and German after his death.

Even though his imaginary geometry was not used directly to describe the world, Lobachevskii found some uses for it in providing geometric interpretations of formulas in analysis. In particular, his paper “Application of imaginary geometry to certain integrals,” which he published in 1836, was translated into German in 1904, with its misprints corrected (Liebmann, 1904). Just as we can compute the seemingly complicated integral

$$\int_0^r \sqrt{r^2 - x^2} dx = \frac{\pi}{4} r^2$$

immediately by recognizing that it represents the area of a quadrant of a circle of radius r , he could use the differential form for the element of area in rectangular coordinates in the

⁷The vast distances between stars make terrestrial units of length inadequate. The light-year (about $9.5 \cdot 10^{12}$ km) is the most familiar unit now used, particularly good, since it tells us “what time it was” when the star emitted the light we are now seeing. Stellar parallax provides another unit, the parsec, which is the distance at which the radius of the earth's orbit subtends an angle of $1''$. A parsec is about 3.258 light-years.

hyperbolic plane given by $dS = (1/\sin y') dx dy$, where y' is the angle of parallelism for the distance y (in our terms $\sin y' = \operatorname{sech} y$) to express certain integrals as the non-Euclidean areas of simple figures. In polar coordinates the corresponding element of area is $dS = \cot r' dr d\theta = \sinh r dr d\theta$. Lobachevskii also gave the elements of volume in rectangular and spherical coordinates and computed 49 integrals representing hyperbolic areas and volumes, including the volumes of pyramids. Using the trigonometry of hyperbolic space, Lobachevskii evaluated a number of integrals, showing, for example, that

$$\int_0^\pi \int_0^\infty 2 \sinh(x) F'(2p \cosh(x) + 2q \cos(\omega) \sinh(x)) dx d\omega = \frac{-\pi F\left(2\sqrt{p^2 - q^2}\right)}{\sqrt{p^2 - q^2}},$$

where $F(x)$ is any continuously differentiable function on $[0, \infty)$ such that $\lim_{x \rightarrow \infty} F(x) = 0$ and $p^2 > q^2$ (Liebmann, 1904, p. 21).

40.6. JÁNOS BÓLYAI

János Bolyai's career turned out less pleasantly than Lobachevskii's. Even though he had the formula for the angle of parallelism in 1823, a time when Lobachevskii was still hoping to vindicate the parallel postulate, he did not publish it until 1831, five years after Lobachevskii's first publication. Even then, he had only the limited space of an appendix to his father's textbook to explain himself. His father sent the appendix to Gauss for comments, and for once Gauss became quite loquacious, explaining that he had had the same ideas many years earlier, and that none of these discoveries were new to him. He praised the genius of the young Bolyai for discovering it, nevertheless. Bolyai the younger was not overjoyed at this response. He suspected Gauss of trying to steal his ideas. According to Paul Stäckel (1862–1919), who wrote the story of the Bolyais, father and son (quoted in Coolidge, 1940, p. 73), when Lobachevskii's work began to be known, Bolyai thought that Gauss was stealing his work and publishing it under the pseudonym Lobachevskii, since "it is hardly likely that two or even three people knowing nothing of one another would produce almost the same result by different routes."

40.7. THE RECEPTION OF NON-EUCLIDEAN GEOMETRY

Some time was required for the new world revealed by Lobachevskii and Bolyai to attract the interest of the mathematical community. Because it seemed possible—even easy—to prove that parallel lines exist or, equivalently, that the sum of the angles of a triangle could not be *more* than two right angles, one can easily understand why a sense of symmetry would lead to a certain stubbornness in attempts to refute the opposite hypothesis as well. Although Gauss had shown the way to a more general understanding with the concept of curvature of a surface (which could be either negative or positive) in the 1825 paper on differential geometry that was published in 1827, it took Riemann's inaugural lecture in 1854 (published in 1867), which made the crucial distinction between the unbounded and the infinite, to give the proper perspective. As Gray (2005, p. 514) says

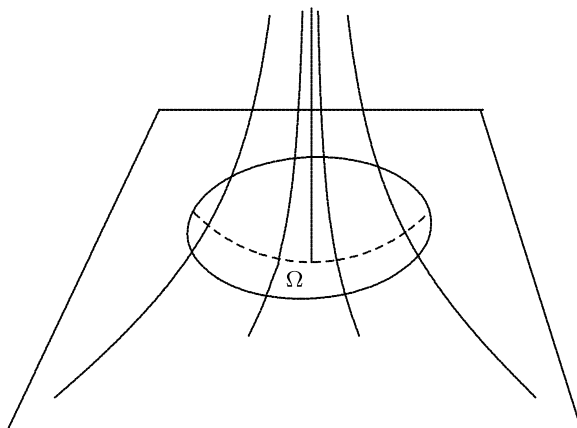


Figure 40.2. Projection of the Lobachevskii-Bolyai plane onto the interior of a Euclidean disk.

[N]ot only was Riemann hostile to the axiomatic treatment of geometry, he was also willing to believe that space was not infinite in extent either.

In 1868, the year after the publication of Riemann's lecture, Beltrami realized that Lobachevskii's theorems provide a model of the Lobachevskii-Bolyai plane in a Euclidean disk. This model is described by Gray (1989, p. 112), as follows. Imagine a directed line perpendicular to the Lobachevskii-Bolyai plane in Lobachevskii-Bolyai three-dimensional space. The entire set of directed lines that are parallel (asymptotic) to this line on the same side of the plane generates a unique horosphere tangent to the plane at its point of intersection with the line. Some of the lines parallel to the given perpendicular in the given direction intersect the original plane, and others do not. Those that do intersect it pass through the portion of the horosphere denoted Ω in Fig. 40.2. Shortest paths on the horosphere are obtained as its intersections with planes passing through the point at infinity that serves as its "center." These paths are called *horocycles*. But there is only one horocycle through a given point in Ω that does not intersect a given horocycle, so that the geometry of Ω is Euclidean. As a result, we have a faithful mapping of the Lobachevskii-Bolyai plane onto the interior of a disk Ω in a Euclidean plane, under which lines in the plane correspond to chords on the disk. This model provides an excellent picture of points at infinity: They correspond to the boundary of the disk Ω . Lines in the plane are parallel if and only if the chords corresponding to them have a common endpoint. Lines that have a common perpendicular in the Lobachevskii-Bolyai plane correspond to chords whose extensions meet outside the circle. It is somewhat complicated to compute the length of a line segment in the Lobachevskii-Bolyai plane from the length of its corresponding chordal segment in Ω or vice versa, and the angle between two intersecting chords is not simply related to the angle between the lines they correspond to.⁸ Nevertheless these computations can be carried out from the trigonometric rules given by Lobachevskii. The result is a perfect model of the Lobachevskii-Bolyai plane *within the Euclidean plane*, obtained by formally reinterpreting

⁸It can be shown that two mutually perpendicular lines correspond to chords having the property that the extension of each passes through the point of intersection of the tangents at the endpoints of the other. But it is far from obvious that this property is symmetric in the two chords, as perpendicularity is for lines.

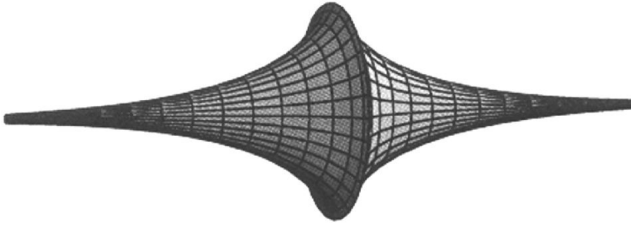


Figure 40.3. The pseudosphere. Observe that it has no definable curvature at its cusp. Elsewhere its curvature is constant and negative.

the words *line*, *plane*, and *angle*. If there were any contradiction in the new geometry, there would be a corresponding contradiction in Euclidean geometry itself.

A variant of this model was later provided by Poincaré, who showed that the diameters and the circular arcs in a disk that meet the boundary in a right angle can be interpreted as lines, and in that case angles can be measured in the ordinary way. Distances are measured using the cross-ratio.

Beltrami also provided a model of a portion of the Lobachevskii–Bólyai plane that could be embedded in three-dimensional Euclidean space: the pseudosphere obtained by revolving a tractrix about its asymptote, as shown in Fig. 40.3.

The pseudosphere is not a model of the entire Lobachevskii–Bólyai plane, since its curvature has a very prominent discontinuity. The problem of finding a surface in three-dimensional Euclidean space that was a perfect model for the Lobachevskii–Bólyai plane, in the sense that its geodesics corresponded to straight lines and lengths and angles were measured in the ordinary way, remained open until Hilbert, in an article “Über Flächen von konstanter Gaußscher Krümmung” (“On surfaces of constant Gaussian curvature”), published in the *Transactions of the American Mathematical Society* in 1901, showed that no such surface exists.

In 1871, Felix Klein gave a discussion of the three kinds of plane geometry in his article “Über die sogenannte nicht-Euklidische Geometrie” (“On the so-called non-Euclidean geometry”), published in the *Mathematische Annalen*. In that article, he gave the classification of them that now stands, saying that the points at infinity on a line were distinct in hyperbolic geometry, imaginary in spherical geometry, and coincident in parabolic (Euclidean) geometry.

40.8. FOUNDATIONS OF GEOMETRY

The problem of the parallel postulate was only one feature of a general effort on the part of mathematicians to improve on the rigor of their predecessors. This problem was particularly acute in the calculus, and the parts of calculus that raised the most doubts were those that were geometric in nature. Euclid, it began to be realized, had taken for granted not only the infinitude of the plane, but also its continuity, and in many cases, had not specified what ordering of points was needed on a line for a particular theorem to be true. If one attempts to prove these theorems without drawing any figures, it becomes obvious what is being assumed. It seemed obvious, for example, that a line joining a point inside a circle to a point outside the circle must intersect the circle in a point, but that fact could not be deduced from Euclid’s axioms. A complete reworking of Euclid was the result, expounded in detail in

Hilbert's *Grundlagen der Geometrie (Foundations of Geometry)*, published in 1903. This book went through many editions and has been translated into English (Bernays, 1971). In Hilbert's exposition, the axioms of geometry are divided into axioms of incidence, order, congruence, parallelism, and continuity, and examples are given to show what cannot be proved when some of the axioms are omitted.

One thing is clear: No new comprehensive geometries are to be expected by pursuing the axiomatic approach of Hilbert. In a way, the geometry of Lobachevskii and Bolyai was a throwback even in its own time. The development of projective and differential geometry would have provided—indeed, *did* provide—non-Euclidean geometry by a natural expansion of the study of surfaces. It was Riemann, not Lobachevskii and Bolyai, who showed the future of geometry. Earlier, we quoted Gray (2005) on Riemann's hostility to the axiomatic Euclidean approach to geometry, often called *synthetic* geometry to distinguish it from analytic geometry, which presumes a metric and the use of numbers to express lengths and areas. Gray also noted (p. 513) that earlier investigators had followed an approach similar to Euclid's, accepting all his axioms except the parallel postulate and then trying to deduce the parallel postulate from the others, an approach that Riemann criticized in his inaugural lecture.

The real "action" in geometry since the early nineteenth century has been in differential, algebraic, and projective geometry. That is not to say that no new theorems can be produced in Euclidean geometry, only that their scope is very limited. There are certainly many such theorems. Coolidge, who undertook the herculean task of writing his *History of Geometric Methods* in 1940, stated in his preface that the subject was too vast to be covered in a single treatise and that "the only way to make any progress is by a rigorous system of exclusion." In his third chapter, on "later elementary geometry," he wrote that "the temptation to run away from the difficulty by not considering elementary geometry after the Greek period at all is almost irresistible." But to attempt to build an entire theory, as Apollonius did, on the synthetic methods and limited techniques in the Euclidean tool kit, would be futile. Even Lobachevskii and Bolyai at least used analytic geometry and trigonometry to produce their results.

PROBLEMS AND QUESTIONS

Mathematical Problems

- 40.1.** Find the Gaussian curvature of the pseudosphere (see Chapter 39 for the definition of Gaussian curvature) obtained by revolving a tractrix about the x -axis. Its parameterization can be taken as

$$\mathbf{r}(u, v) = \left(a \operatorname{sech}\left(\frac{u}{a}\right) \cos(v), a \operatorname{sech}\left(\frac{u}{a}\right) \sin(v), u - a \tanh\left(\frac{u}{a}\right) \right).$$

(The parameter u ranges over all real numbers, v over an interval of length 2π .) Explain why the pseudosphere can be thought of as "a sphere of imaginary radius."

- 40.2.** Consider the two Pythagorean theorems in elliptic and hyperbolic geometry exhibited above, and assume the radius of curvature r is 1 in both cases. How long is the hypotenuse of an isosceles right triangle whose legs are each of length $3/2$? Compare

those lengths with the Euclidean length. Which of the three is smallest, and which largest?

- 40.3.** Suppose two chords on a circle, say AB and CD are such that the tangents to the circle at A and B meet at a point on the extension of CD . (This means, in the Beltrami model of hyperbolic geometry, that AB is perpendicular to CD .) Prove that the tangents to the circle at C and D meet at a point on the extension of AB . In other words, if AB is perpendicular to CD , then CD is perpendicular to AB .

Historical Questions

- 40.4.** In what ways did Saccheri duplicate unknowingly earlier work by Thabit ibn-Qurra. (See Chapter 27.)
- 40.5.** Why did Gauss not publish his research in hyperbolic geometry?
- 40.6.** What considerations finally led to the recognition by all practicing mathematicians that non-Euclidean geometry was consistent?

Questions for Reflection

- 40.7.** Gauss realized that the geometry of physical space could be determined by observation, measuring the angle sum of very large triangles. (Theory shows that if space is homogeneous and non-Euclidean, then the larger a triangle is, the more its angle sum will differ from two right angles.) While he was involved in the survey of Hannover, he tried to determine the angles of a large triangle formed by three mountaintops. The three that he used, however, were not far enough apart to show any significant deviation from two right angles. If we use a radius or diameter of the earth's orbit as one side of the triangle and a nearby fixed star as the opposite vertex, the result is a truly large triangle. Such measurements became feasible during the early nineteenth century, and the two angles that can be measured from the earth (at opposite ends of the radius or diameter turned out to be very nearly, but not quite, two right angles. What is the proper conclusion? (1) Space is Euclidean, and the angle at the star (its parallax) is the supplement of the sum of these two angles? (2) Space is hyperbolic, and the angle at the star is smaller than that supplement? (3) Space is elliptic, and the angle at the star is larger than that supplement? Can these alternatives be distinguished by any observation from the earth?
- 40.8.** Consistency and applicability are two very different issues in the world of mathematics. Granted that the consistency of non-Euclidean geometry was accepted by all mathematicians by the end of the nineteenth century, what applications have been found for these new geometries?
- 40.9.** A rear-guard battle against enlightenment can be maintained for a surprisingly long time, even by people who have some scientific competence and even after there is a general consensus as to the truth. The long history of circle squarers and angle trisectors is a good example of this phenomenon. Most of those who work on such problems are non-mathematicians who simply don't understand the meaning of infinite precision. They waste their time, but can be made to recognize that a particular effort has failed, even as they turn again with renewed vigor to engage in this

hopeless enterprise. Only a few of them are so logic-impaired that they are completely incapable of coherent reasoning. In the case of non-Euclidean geometry, this mathematical aberration sometimes expresses itself in attempts to prove the parallel postulate. One such attempt, from 80 years ago, was made by a quite intelligent and mathematically semi-literate scholar, the Rev. Dr. Jeremiah J. Callahan, president (at the time) of Duquesne University. Dr. Callahan wrote a treatise entitled *Euclid or Einstein*, in which he “proved” the parallel postulate by redefining parallel lines as “lines that are equidistant at equidistant points,” not realizing that the assumption that such lines exist is equivalent to the parallel postulate. He thought he had a proof that his definition was equivalent to Euclid’s definition. One can have some sympathy for him on this point, since Euclid’s definition of a line is “that which lies evenly along itself,” hardly a happy effort, since it would apply equally well to a circle.⁹

These “mathematical cranks” very frequently attempt to publish their work in newspapers or get them reported as news. What should mathematicians do when confronted by reporters asking them to comment on such work? Which is the better strategy: patient explanation or open contempt? Should the goal be to bring the crank to recognize his (it’s almost always *his*, not *her*) errors? Or should it be to make the public laugh at the crank? Or to get the public to understand the proper relation between science and the nonspecialist citizen?

⁹The Greek word for a straight line is *eutheîa*, from *eu-* (good, well) and the root *the-* (put, set, as in our loan word *thesis*).

Complex Analysis

In the mid-1960s, the late Walter Rudin (1921–2010), the author of several standard graduate textbooks in mathematics, wrote a textbook with the title *Real and Complex Analysis*, aimed at showing the considerable unity and overlap between the two subjects. It was necessary to write such a book because real and complex analysis, while sharing common roots in the calculus, had developed quite differently. The contrasts between the two are considerable. Complex analysis considers the smoothest, most orderly possible functions, those that are analytic, while real analysis allows the most chaotic imaginable functions. Complex analysis was, to pursue our botanical analogy, fully a “branch” of calculus, and foundational questions hardly entered into it. Real analysis had a share in both roots and branches, and it was intimately involved in the debate over the foundations of calculus.

What caused the two varieties of analysis to become so different? Both deal with functions, and both evolved under the stimulus of the differential equations of mathematical physics. The central point is the concept of a function. We have already seen the early definitions of this concept by Leibniz and John Bernoulli. All mathematicians from the seventeenth and eighteenth centuries had an intuitive picture of a function as a formula or expression in which variables are connected by rules derived from algebra or geometry. A function was regarded as continuous if it was given by a single formula throughout its range. If the formula changed, the function was called “mechanical” by Euler. Although “mechanical” functions may be continuous in the modern sense, they are not usually analytic. All the “continuous” functions in the older sense are analytic. They have power-series expansions, and those power-series expansions are often sufficient to solve differential equations. As a general signpost indicating where the paths diverge, the path of power-series expansions and the path of trigonometric-series expansions is a rough guide. A consequence of the development was that real-variable theory had to deal with very irregular and “badly behaved” functions. It was therefore in real analysis that the delicate foundational questions arose. This chapter and the two following are devoted to exploring these developments.

41.1. IMAGINARY AND COMPLEX NUMBERS

Although imaginary numbers seem more abstract to modern mathematicians than negative and irrational numbers, that is because their physical interpretation is more remote from

everyday experience. One interpretation of $i = \sqrt{-1}$, for example, is as a rotation through a right angle. (Since $i^2 = -1$, the effect of multiplying a real number twice by i is to rotate the real axis by 180° . Hence multiplying once by i ought to rotate it by 90° .)

We have an intuitive picture of the length of a line segment and decimal approximations to describe that length as a number; that is what gives us confidence that irrational numbers really are numbers. But it is difficult to think of a rotation as a number. On the other hand, the rules for multiplying complex numbers—at least those whose real and imaginary parts are rational—are much simpler and easier to understand than the definition of irrational numbers. In fact, complex numbers were understood before real numbers were properly defined; mathematicians began trying to make sense of them as soon as there was a clear need to do so. That need came not, as one might expect, from trying to solve quadratic equations such as $x^2 - 2x + 2 = 0$, where the quadratic formula produces $x = 1 \pm \sqrt{-1}$. It was possible in this case simply to say that the equation had no solution.

To find the origin of imaginary numbers, we need to return to the algebraic solution of cubic equations that we discussed in Chapters 30 and 37. Recall that the algorithm for finding the solution had the peculiar property that it involved taking the square root of a negative number precisely when there were three distinct real solutions. For example, the algorithm gives the solution of $x^3 - 7x + 6 = 0$ as

$$x = \sqrt[3]{3 - \sqrt{-\frac{100}{27}}} - \sqrt[3]{3 + \sqrt{-\frac{100}{27}}}.$$

We cannot say that the equation has no roots, since it obviously has 1, 2, and -3 as roots. Thus the challenge arose: Make sense of this formula. Make it say “1, 2, and -3 .”

This challenge was taken up by Bombelli, who wrote a treatise on algebra which he wrote in 1560, but which was not published until 1572. In that treatise he invented the name “plus of minus” to denote a square root of -1 and “minus of minus” for its negative. He did not think of these two concepts as different numbers, but rather as the *same* number being added in the first case and subtracted in the second. What is most important is that he realized what rules must apply to them in computation: plus of minus times plus of minus makes minus and minus of minus times minus of minus makes minus, while plus of minus times minus of minus makes plus. Bombelli had an *ad hoc* method of taking the cube root of a complex number, opportunistically taking advantage of any extra symmetry in the number whose root was to be extracted. In considering the equation $x^3 = 15x + 4$, for example, he found by applying the formula that $x = \sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{2 - \sqrt{-121}}$. In this case, however, Bombelli was able to work backward, since he knew in advance that one root is 4; the problem was to make the formula say “4.” Bombelli had the idea that the two cube roots must consist of real numbers together with his “plus of minus” or “minus of minus.” Since the numbers under the cube root sign are (as we would say) complex conjugates of each other, it would seem likely that the two cube roots are as well. That is the real parts are equal, and the imaginary parts are negatives of each other. Since the sum of the two cube roots is four, it follows that the real parts must be 2. Thus $\sqrt[3]{2 \pm 11\sqrt{-1}} = 2 \pm t\sqrt{-1}$. The number t is now easily found by cubing: $2 \pm 11\sqrt{-1} = (8 - 6t^2) \pm (12t - t^3)\sqrt{-1}$. Obviously, $t = 1$, and so $\sqrt[3]{1 \pm 11\sqrt{-1}} = 2 \pm \sqrt{-1}$. If we didn’t know a root, this approach would lead nowhere; but if a solution is given, it explains how the imaginary numbers are to be interpreted and used in computation.

41.1.1. Wallis

In an attempt to make these numbers more familiar, the English mathematician John Wallis (1616–1703) pointed out that while no positive or negative number could have a negative square, nevertheless it is also true that no physical quantity can be negative, that is, less than nothing. Yet negative numbers were accepted and interpreted as retreats when the numbers measure advances along a line. Wallis thought that what was allowed in lines might also apply to areas, pointing out that if 30 acres are reclaimed from the sea, and 40 acres are flooded, the net amount “gained” from the sea would be -10 acres. Although he did not say so, it appears that he regarded this real loss of 10 acres as an imaginary gain of a square of land $\sqrt{-435600} = 660\sqrt{-1}$ feet on a side.

What he did say in his 1673 treatise on algebra was that one could represent $\sqrt{-ab}$ as the mean proportional between a and $-b$. The mean proportional is easily found for two positive line segments a and b . Simply lay them end to end, use the union as the diameter of a circle, and draw the half-chord perpendicular to that diameter at the point where the two segments meet. That half-chord (sine) is the mean proportional. If only mathematicians had used the Euclidean construction of the mean proportional, interpreting points to the left of 0 as negative, they would have gotten the geometric interpretation of imaginary numbers as points on an axis perpendicular to the real numbers, as shown in the drawing on the left in Fig. 41.1.

As things turned out, however, this idea took some time to catch on. Wallis’ thinking went in a different direction. When one of the numbers was regarded as negative, he regarded the negative quantity as an oppositely directed line segment. He then modified the construction of the mean proportional between the two segments. When two oppositely directed line segments are joined end to end, one end of the shorter segment lies between the point where the two segments meet and the other end of the longer segment, so that the point where the segments join up lies *outside* the circle passing through their other two endpoints. Wallis interpreted the mean proportional as the tangent to the circle from the point where the two segments meet. Thus, whereas the mean proportional between two positive quantities is represented as a sine, that between a positive and negative quantity is represented as a tangent. This approach is quite consistent, since the endpoint of b can move in either direction without upsetting the numerical relationship between the lines. And indeed, it is easy to verify that the *length* of Wallis’ tangent line is indeed the mean proportional between the lengths of the two given lines.

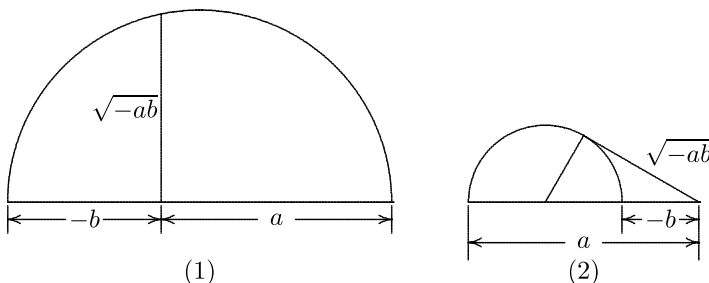


Figure 41.1. (1) How it might have been: The mean proportional between a positive number a and a negative number $-b$ lies on an axis perpendicular to them. (2) How Wallis thought of it: The mean proportional is a tangent instead of a sine.

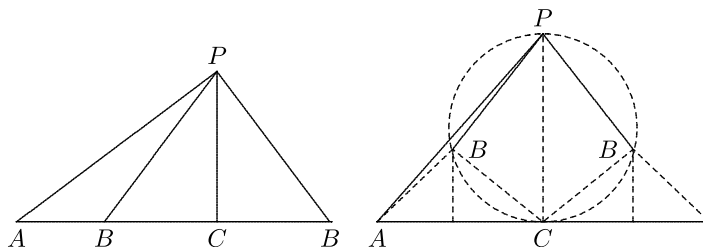


Figure 41.2. Wallis' geometric solution of quadratic equation with real roots (left) and complex roots (right).

Wallis applied this procedure in an “imaginary” construction problem. First he stated the following “real” problem. Given a triangle having side AP of length 20, side PB of length 15, and altitude PC of length 12, find the length of side AB , taken as base in Fig. 41.2. Wallis pointed out that two solutions were possible. Using the foot of the altitude as the reference point C and applying the Pythagorean theorem twice, he found that the possible lengths of AB were 16 ± 9 , that is, 7 and 25. (In general $PB = \sqrt{AP^2 - PC^2} \pm \sqrt{PB^2 - PC^2}$.) He then proposed reversing the data, in effect considering an impossible triangle having side AP of length 20, side PB of length 12, and altitude PC of length 15. The two solutions would thus appear to be $5\sqrt{7} \pm 9\sqrt{-1}$. Although the algebraic problem has no real solution, a fact verified by the geometric figure (Fig. 41.1), one could certainly draw the two line segments AB . These line segments could therefore be interpreted as the numerical solutions of the equation, representing a triangle with one side (AB) having imaginary (actually complex) length: $AB = AC \pm BC$. In the figure $AC = 5\sqrt{7}$, and so the two sides drawn as BC (which really do have length 9) represent $\pm 9\sqrt{-1}$. Once again, the imaginary part of the representation is oblique to the real part, but not perpendicular to it.

The rules given by Bombelli had made imaginary and complex numbers accessible, and they turned out to be very convenient in many formulas. Euler made free use of them, studying power series in which the variables were allowed to be complex numbers and deriving a famous formula

$$e^{v\sqrt{-1}} = \cos v + \sqrt{-1} \sin v.$$

Euler derived this result in a paper on ballistics written around 1727 (see Smith, 1929, pp. 95–98), just after he moved to Russia. He had no thought of representing $\sqrt{-1}$ as we now do, on a line perpendicular to the real axis.

41.1.2. Wessel

Wallis' work had given the first indication that complex numbers would have to be interpreted as line segments in a plane, a discovery made again 70 years later by the Norwegian surveyor Caspar Wessel (1745–1818). The only mathematical paper he ever wrote was delivered to the Royal Academy in Copenhagen, Denmark in 1797, but he had been in possession of the results for about a decade at that time. In that paper (Smith, 1929, pp. 55–66), he explained how to multiply lines in a plane by multiplying their lengths and adding the angles they make with a given reference line, on which a length is chosen to represent +1:

Let $+1$ designate the positive rectilinear unit and $+\epsilon$ a certain other unit perpendicular to the positive unit and having the same origin; the direction angle of $+1$ will be equal to 0° , that of -1 to 180° , that of $+\epsilon$ to 90° , and that of $-\epsilon$ to -90° or 270° . By the rule that the direction angle of the product shall equal the sum of the angles of the factors, we have: $(+1)(+1) = +1$; $(+1)(-1) = -1$; $(-1)(-1) = +1$; $(+1)(+\epsilon) = +\epsilon$; $(+1)(-\epsilon) = -\epsilon$; $(-1)(+\epsilon) = -\epsilon$; $(-1)(-\epsilon) = +\epsilon$; $(+\epsilon)(+\epsilon) = -1$; $(+\epsilon)(-\epsilon) = +1$; $(-\epsilon)(-\epsilon) = -1$. From this it is seen that ϵ is equal to $\sqrt{-1}$. [Smith, 1929, p. 60]

Wessel noticed the connection of these rules with the addition and subtraction formulas for sign and cosine and gave the formula $(\cos x + \epsilon \sin x)(\cos y + \epsilon \sin y) = \cos(x + y) + \epsilon \sin(x + y)$. On that basis he was able to reduce the extraction of the n th root of a complex number to extracting the same root for a positive real number and dividing the polar angle by n .

41.1.3. Argand

The reaction of the mathematical community to this simple but profound idea was less than overwhelming. Wessel's work was forgotten for a full century. In the meantime, another mathematician by avocation, the French accountant Jean Argand (1768–1822), published the small book *Essai sur une manière de représenter les quantités imaginaires dans les constructions géométriques* at his own expense in 1806, modestly omitting to name himself as its author, in which he advocated essentially the same idea, thinking, as Wallis had done, of an imaginary number as the mean proportional between a positive number and a negative number. Through a complicated series of events this book and its author gradually became known in the mathematical community.

There was resistance to the idea of interpreting complex numbers geometrically, since they had arisen in algebra. Geometry was essential to the algebra of complex numbers, as shown by the fact that a proof of the fundamental theorem of algebra by Gauss in 1799 is based on the idea of intersecting curves in a plane. The lemmas that Gauss used for the proof had been proved earlier by Euler using the algebra of imaginary numbers, but Gauss gave a new proof using only real numbers, precisely to avoid invoking any properties of imaginary numbers. Nevertheless, because he developed a good portion of the theory of complex integrals and analytic function theory, the complex plane is now often called the *Gaussian plane*.¹

Even though he avoided the algebra of imaginary numbers, Gauss still needed the continuity properties of real numbers, which, as we just saw, were not fully arithmetized until many years later.² Continuity was a geometric property that occurred implicitly in Euclid, but Gauss expressed the opinion that continuity could be arithmetized. In giving a fifth proof of this theorem half a century later, he made full use of complex numbers.

¹Hille (1959, p. 18) noted that the representation of complex numbers on the plane is called *le diagramme d'Argand* in France and credited the Norwegians with "becoming modesty" for not claiming *det Wesselske planet*.

²The Czech scholar Bernard Bolzano (1781–1848) showed how to approach the idea of continuity analytically in an 1817 paper. His work anticipated Dedekind's arithmetization of real numbers, which will be discussed in the next chapter.

41.2. ANALYTIC FUNCTION THEORY

Calculus began with a limited stock of geometry: a few curves and surfaces, all of which could be described analytically in terms of rational, trigonometric, exponential, and logarithmic functions of real variables. Soon, however, calculus was used to formulate problems in mathematical physics as differential equations. To solve those equations, the preferred technique was integration, but where integration failed, power series were the technique of first resort. These series automatically brought with them the potential of allowing the variables to assume complex values, since a series expansion in powers of $x - x_0$ that converges at $x = x_1$, automatically converges for all complex numbers inside the circle through x_1 with center at x_0 . But then, integration and differentiation had to be defined for complex functions of a complex variable. The result was a theory of analytic functions of a complex variable involving complex integrals. The scope of this theory was much vaster than the materials that led to its creation.

In his 1748 *Introductio*, Euler emended the definition of a function, saying that a function is an *analytic expression* formed from a variable and constants. The rules for manipulating symbols were agreed on as long as only finite expressions were involved. But what did the symbols *represent*? Euler stated that variables were allowed to take on negative and imaginary values. Thus, even though the physical quantities the variables represented were measured as *positive rational* numbers, the algebraic and geometric properties of negative, irrational, and complex numbers could be invoked in the analysis. The extension from finite to infinite expressions was not long in coming.

Lagrange undertook to reformulate the calculus in his treatises *Théorie des fonctions analytiques* (1797) and *Leçons sur le calcul des fonctions* (1801), basing it entirely on algebraic principles and stating as a fundamental premise that the functions to be considered are those that can be expanded in power series (having no negative or fractional powers of the variable). With this approach, the derivatives of a function need not be defined as ratios of infinitesimals, since they can be defined in terms of the coefficients of the series that represents the function. Functions having a power series representation are known nowadays as *analytic functions* from the title of Lagrange's work.

41.2.1. Algebraic Integrals

Early steps toward complexification were taken only on a basis of immediate necessity. As we have already seen, the applications of calculus in solving differential equations made the computation of integrals extremely important. Now, computing the derivative never leads outside the class of elementary functions and leaves algebraic functions algebraic, trigonometric functions trigonometric, and exponential functions exponential; integrals, however, are a very different matter. Algebraic functions often have nonalgebraic integrals, as Leibniz realized very early. The relation we now write as

$$\arccos(1 - x) = \int_0^x \frac{1}{\sqrt{2t - t^2}} dt$$

was written by him as

$$a = \int dx : \sqrt{2x - x^2},$$

where $x = 1 - \cos a$. Eighteenth-century mathematicians were greatly helped in handling integrals like this by the use of trigonometric functions. It was therefore natural that they would see the analogy when more complicated integrals came to be considered. Such problems arose from the study of pendulum motion and the rotation of solid bodies in physics, but we shall illustrate it with examples from pure geometry: the rectification of the ellipse and the division of the lemniscate³ into equal arcs. For the circle, we know that the corresponding problems lead to the integral

$$\int_0^x \frac{1}{\sqrt{1-t^2}} dt$$

for the length of the arc of the unit circle above the interval $[0, x]$ and an equation

$$\int_0^y \frac{1}{\sqrt{1-t^2}} dt = \frac{1}{n} \int_0^x \frac{1}{\sqrt{1-t^2}} dt,$$

which can be written in differential form as

$$\frac{dx}{\sqrt{1-x^2}} = \frac{n dy}{\sqrt{1-y^2}},$$

for the division of that arc into n equal pieces.

Trigonometry helps to solve this last equation. Instead of the function $\arcsin(x)$ or $-\arccos(x)$ that the integral actually represents, it makes more sense to look at an inverse of it, say the cosine function. This function provides an algebraic equation through its addition formula,

$$a_0 y^n - a_2 y^{n-2} + a_3 y^{n-4} - \dots = x,$$

relating the abscissas of the end of the given arc (x) and the end of the n th part of it (y). The algebraic nature of this equation determines whether the division problem can be solved with ruler and compass. In particular, for $n = 3$ and a 60-degree arc ($x = 1/2$), for which the equation is $4y^3 - 3y = 1/2$, such a solution does not exist. Thus the problems of computing arc length on a circle and equal division of its arcs lead to an interesting combination of algebra, geometry, and calculus. Moreover, the periodicity of the inverse function of the integral helps to find all solutions of this equation.

The division problem was fated to play an important role in study of integrals of algebraic functions. The Italian nobleman Giulio de' Toschi Fagnano (1682–1766) studied the problem of rectifying the lemniscate, whose polar equation is $r^2 = 2 \cos(2\theta)$. Its element of arc is $\sqrt{2}(1 - 2 \sin^2 \theta)^{-1/2} d\theta$, and the substitution $u = \tan \theta$ turns this expression into $\sqrt{2}(1 - u^4)^{-1/2} du$. Thus, the rectification problem involves evaluating the integral

$$\int_0^x \frac{\sqrt{2}}{\sqrt{1-u^4}} du,$$

³The simplest lemniscate, first described by James Bernoulli in 1694, is the set of points the product of whose distances from the points $(+a, 0)$ and $(-a, 0)$ is a^2 . It looks like a figure eight and has equation $(x^2 + y^2)^2 = 2a^2(x^2 - y^2)$.

while the division problem involves solving the differential equation

$$\frac{dz}{\sqrt{1-z^4}} = \frac{n du}{\sqrt{1-u^4}}.$$

Fagnano gave the solution for $n = 2$ as the algebraic relation

$$\frac{u\sqrt{2}}{\sqrt{1-u^4}} = \frac{1}{z} \sqrt{1 - \sqrt{1-z^4}}.$$

Euler observed the analogy between these integrals and the circular integrals just discussed, and suggested that it would be reasonable to study the inverse function. But Euler lived at a time when the familiar functions were still the elementary ones. He found a large number of integrals that could be expressed in terms of algebraic, logarithmic, and trigonometric functions and showed that there were others that could not be so expressed.

41.2.2. Legendre, Jacobi, and Abel

The foundation for further work in integration was laid by Legendre, who invented the term *elliptic integral*. Off and on for some 40 years between 1788 and 1828, he thought about integrals like those of Fagnano and Euler, classified them, computed their values, and studied their properties. He found their algebraic addition formulas and thereby reduced the division problem for these integrals to the solution of algebraic equations. Interestingly, he found that whereas the division problem requires solving an equation of degree n for the circle, it requires solving an equation of degree n^2 for the ellipse. After publishing his results as exercises in integral calculus in 1811, he wrote a comprehensive treatise in the 1820s. As he was finishing the third volume of this treatise he discovered a new set of transformations of elliptic integrals that made their computation easier. (He already knew one set of such transformations.) Just after the treatise appeared in 1827, he found to his astonishment that Jacobi had discovered the same transformations, along with others, and had connected them with the division problem. Jacobi's results in turn were partially duplicated by those of Abel.

Abel, who admired Gauss, was proud of having achieved the division of the lemniscate into 17 equal parts,⁴ just as Gauss had done for the circle. The secret for the circle was to use the algebraic addition formula for trigonometric functions. For the lemniscate, as Legendre had shown, the equation was of higher degree. Abel was able to solve it by using complex variables, and in the process, he discovered that the inverse functions of the elliptic integrals, when regarded as functions of a complex variable, were *doubly periodic*. The double period accounted for the fact that the division equation was of degree n^2 rather than n . Without complex variables, the theory of elliptic integrals would have been a disconnected collection of particular results. With them, a great simplicity and unity was achieved. Abel went on to study algebraic addition formulas for very general integrals of the type

$$\int R(x, y(x)) dx,$$

⁴Or, more generally, a Fermat prime number of parts.

where $R(x, y)$ is a rational function of x and y and $y(x)$ satisfies a polynomial equation $P(x, y(x)) = 0$. Such integrals are now called *abelian integrals* in his honor. In so doing, he produced one of the ground-breaking theorems of the early nineteenth century.

The model for Abel's theorem, as for so much of algebraic function theory, comes from the theory of trigonometric functions. The fact that

$$\int_0^x \frac{1}{\sqrt{1-t^2}} dt = \arcsin(x)$$

combines with the addition law $\sin(u+v) = \sin(u)\cos(v) + \sin(v)\cos(u)$ to produce an addition law for these integrals.

$$\int_0^x \frac{1}{\sqrt{1-t^2}} dt + \int_0^y \frac{1}{\sqrt{1-t^2}} dt = \int_0^z \frac{1}{\sqrt{1-t^2}} dt,$$

where $z = x\sqrt{1-y^2} + y\sqrt{1-x^2}$. By induction, the sum of any number of such integrals can be reduced to a single integral whose upper limit is an algebraic function of the upper limits of the terms in the sum. In particular, because z is an algebraic function of x and y the addition formula reduces the problem of trisecting an angle to a matter of solving the equation $4u^3 - 3u = v$, where $v = \cos(\theta)$ and $u = \cos(\theta/3)$. In general, dividing an arc into n equal pieces is a matter of solving an equation of degree n .

As Legendre discovered, the same is true for the elliptic integral:

$$\int_0^x \frac{1}{\sqrt{(1-t^2)(1-c^2t^2)}} dt + \int_0^y \frac{1}{\sqrt{(1-t^2)(1-c^2t^2)}} dt = \int_0^z \frac{1}{\sqrt{(1-t^2)(1-c^2t^2)}},$$

where

$$z = \frac{x\sqrt{(1-y^2)(1-c^2y^2)} + y\sqrt{(1-x^2)(1-c^2x^2)}}{1-c^2x^2y^2}.$$

Thus, the sum of any number of elliptic integrals can be reduced to a single integral whose upper limit is an algebraic function of their upper limits. Again, this means that dividing an arc of the lemniscate into n equal pieces is a matter of solving an algebraic equation. However, because of the complexity of the addition formula for elliptic integrals, that equation is of degree n^2 , not n .

Abel established a great generalization of this fact: For each polynomial $P(x, y)$, there is a number p , now called the *genus* of the curve $P(x, y) = 0$, such that a sum of any number of integrals having $R(x, y)$ as an integrand with different limits of integration can be expressed in terms of just p integrals, whose limits of integration are algebraic functions of those in the given sum.⁵ For elliptic integrals, $p = 1$, and that is the content of the algebraic addition formulas discovered by Legendre. For a more complicated integral, say

$$\int \frac{1}{\sqrt{q(x)}} dx,$$

⁵To avoid complications, we are not discussing Legendre's three kinds of elliptic integrals. For those who know what they are, the results we state here should be assumed to apply only to integrals of first kind.

where $q(x)$ is a polynomial of degree 5 or higher, the genus will be higher. If $P(x, y) = y^2 - q(x)$, where the polynomial q is of degree $2p + 1$ or $2p + 2$, the genus is p .

After Abel's premature death in 1829, Jacobi continued to develop algebraic function theory. In 1832, he realized that for algebraic integrals of higher genus, the limits of integration in the p integrals to which a sum was reduced could not be determined, since there were p integrals and only one equation connecting them to the variable in terms of which they were to be expressed. He therefore had the idea of adjoining extra equations in order to determine these limits. For example, if $q(x)$ is of degree 5, he posed the problem of solving for x and y in terms of u and v in the equations

$$u = \int_0^x \frac{1}{\sqrt{q(t)}} dt + \int_0^y \frac{1}{\sqrt{q(t)}} dt$$

$$v = \int_0^x \frac{t}{\sqrt{q(t)}} dt + \int_0^y \frac{t}{\sqrt{q(t)}} dt.$$

This problem became known as the *Jacobi inversion problem*. Solving it took a quarter of a century and led to progress in both complex analysis and algebra.

41.2.3. Theta Functions

Jacobi himself gave this solution a start in connection with elliptic integrals. Although a nonconstant function that is analytic in the whole plane cannot be doubly periodic (because its absolute value cannot attain a maximum), a quotient of such functions can be, and Jacobi found the ideal numerators and denominators to use for expressing the doubly periodic elliptic functions as quotients: theta functions. The secret of solving the Jacobi inversion problem was to use theta functions in more than one complex variable, but working out the proper definition of those functions and the mechanics of applying them to the problem required the genius of Riemann and Weierstrass. These two giants of nineteenth-century mathematics solved the problem independently and simultaneously in 1856, but considerable preparatory work had been done in the meantime by other mathematicians. The importance of algebraic functions as the basic core of analytic function theory cannot be overemphasized. Klein (1926, p. 280) goes so far as to say that Weierstrass' purpose in life was

to conquer the inversion problem, even for hyperelliptic integrals of arbitrarily high order, as Jacobi had foresightedly posed it, perhaps even the problem for general abelian integrals, using rigorous, methodical work with power series (including series in several variables). It was in this way that the topic called the Weierstrass theory of analytic functions arose as a by-product.

41.2.4. Cauchy

Cauchy's name is associated most especially with one particular approach to the study of analytic functions of a complex variable, that based on complex integration. A complex variable is really two variables, as Cauchy was saying even as late as 1821. But a function is to be given by the same symbols, whether they denote real or complex numbers. When we integrate and differentiate a given function, which variable should we use? Cauchy discovered the answer, as early as 1814, when he first discussed such questions in print.

The value of the function is a complex number that can also be represented in terms of two real numbers u and v , as $u + iv$. If the derivative is to be independent of the real variable on which it is taken, these must satisfy the equations we now call the Cauchy–Riemann equations:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}; \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

In that case, as Cauchy saw, if we are integrating $u + iv$ in a purely formal way, separating real and imaginary parts, over a path from the lower left corner of a rectangle (x_0, y_0) to its upper right corner (x_1, y_1) , the same result is obtained whether the integration proceeds first vertically, then horizontally or first horizontally, then vertically. As Gauss had noted as early as 1811, Cauchy observed that the function $1/(x + iy)$ did not have this property if the rectangle contained the point $(0, 0)$. The difference between the two integrals in this case was $2\pi i$, which Cauchy called the *residue*. Over the period from 1825 to 1840, Cauchy developed from this theorem what is now known as the Cauchy integral theorem, the Cauchy integral formula, Taylor’s theorem, and the calculus of residues. The Cauchy integral theorem states that if γ is a closed curve inside a simply connected region⁶ in which $f(z)$ has a derivative then

$$\int_{\gamma} f(z) dz = 0.$$

If the real and imaginary parts of this integral are written out and compared with the Cauchy–Riemann equations, this formula becomes a simple consequence of what is known as Green’s theorem (the two-variable version of the divergence theorem), published in 1828 by George Green (1793–1841) and simultaneously in Russia by Mikhail Vasilevich Ostrogradskii (1801–1862). When combined with the fact that the integral of $1/z$ over a curve that winds once around 0 is $2\pi i$, this theorem immediately yields as a consequence the Cauchy integral formula

$$f(z_0) = \frac{1}{2\pi i} \int_{\gamma} \frac{f(z)}{z - z_0} dz.$$

When generalized, this formula becomes the residue theorem. Also from it, one can obtain estimates for the size of the derivatives. Finally, by expanding the denominator as a geometric series in powers of $z - z_1$, where z_1 lies inside the curve γ , one can obtain the Taylor series expansion of $f(z)$. These theorems form the essential core of modern first courses in complex analysis. This work was supplemented by a paper of Pierre Laurent (1813–1854), submitted to the Paris Academy in 1843, in which power series expansions about isolated singularities (Laurent series) were studied.

Cauchy was aware of the difficulties that arise in the case of multivalued functions and introduced the idea of a barricade (*ligne d’arrêt*) to prevent a function from assuming more than one value at a given point. As mentioned in Section 38.3 of Chapter 38, his student

⁶See Chapter 38 for the definition of a simply connected region.

Puiseux studied the behavior of algebraic functions in the neighborhood of what we now call branch points, which are points where the distinct values of a multi-valued function coalesce two or more at a time.

41.2.5. Riemann

The work of Puiseux on algebraic functions of a complex variable was to be subsumed in two major papers of Riemann. The first of these, his doctoral dissertation, contained the concept now known as a Riemann surface. It was aimed especially at simplifying the study of an algebraic function $w(z)$ satisfying a polynomial equation $P(z, w(z)) \equiv 0$. In a sense, the Riemann surface revealed that all the significant information about the function was contained precisely in its singularities—the way it branched at its branch points. Information about the surface was contained in its *genus*, defined as half the total number of branch points, counted according to order, less the number of sheets in the surface, plus 1.⁷ The Riemann surface of $w = \sqrt{z}$, for example, has two branch points (0 and ∞), each of order 1, and two sheets, resulting in genus 0.

The smooth transition from z_1 to z_2 is shown in Fig. 41.3.⁸ The two z -planes are cut along the positive real axis or any other ray emanating from the “branch point” 0. Then the lower side of each cut is glued to the upper side of the other. (This will be easiest to visualize if you imagine the z_2 -plane picked up and turned over so that the dotted edge of the z_2 -plane lies on the dotted edge of the z_1 plane.) The result is the *Riemann surface* of the function $w = \sqrt{z}$. It consists of two “sheets” (copies of the complex plane) glued together as just stated. You can easily make a model of this surface with two sheets of paper, a pair of scissors, and cellophane tape. On such a model you can move your finger smoothly and continuously over the entire Riemann surface, without any jumps when it moves from the z_1 sheet to the z_2 sheet. In particular, if you describe a small circle about the branch point 0 at the end of the cut, you will see that it crosses over to the back of the paper when it moves across the dotted edges that have been glued together, makes a whole circle on the back, then crosses over again to the front when it moves across the solid line.

At every point on the Riemann surface except the branch point $z = 0$, the mapping $z \mapsto w$ is *analytic*; that is, it has a power-series representation. For example, near the point $z_1 = 2$, we can express w as a series of powers of $z - 2$ using the binomial theorem:

$$\begin{aligned}\sqrt{z} &= \sqrt{2 + z - 2} = \sqrt{2}\sqrt{1 + (z - 2)/2} \\ &= \sqrt{2}\left(1 + \frac{1}{2}(z - 2) + \frac{\frac{1}{2}(-\frac{1}{2})}{2^2 2!}(z - 2)^2 + \frac{\frac{1}{2}(-\frac{1}{2})(-\frac{3}{2})}{2^3 3!}(z - 2)^3 + \dots\right) \\ &= \sqrt{2}\left(1 + \frac{1}{2}(z - 2) - \frac{1}{32}(z - 2)^2 + \frac{1}{128}(z - 2)^3 - \frac{5}{2048}(z - 2)^4 + \dots\right).\end{aligned}$$

Riemann’s geometric approach to the subject brought out the duality between surfaces and mappings of them, encapsulated in a formula known as the Riemann–Roch theorem (after Gustav Roch, 1839–1866). This formula connects the dimension of the space of functions on a Riemann surface having prescribed zeros and poles (places where it becomes infinite) with the genus of the surface.

⁷Klein (1926, p. 258) ascribes this definition to Alfred Clebsch (1833–1872).

⁸This figure is taken from the author’s *Classical Algebra* (Wiley, 2008).

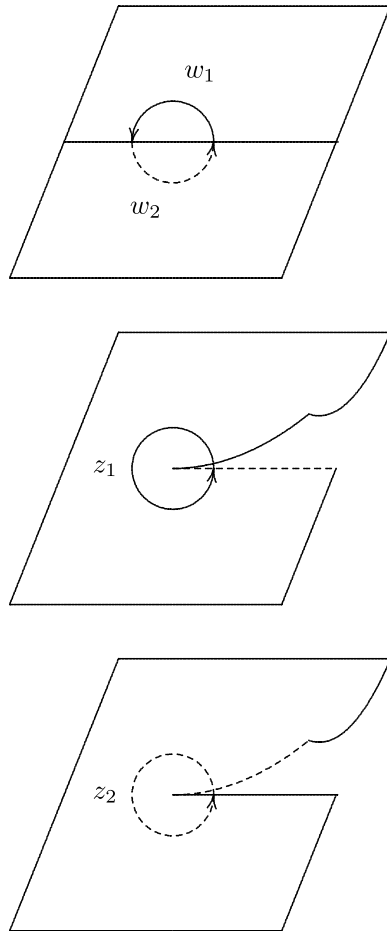


Figure 41.3. The Riemann surface of $w = \sqrt{z}$. The solid z_1 circle corresponds to the solid w_1 semicircle, and the dashed z_2 circle to the dashed w_2 semicircle.

In 1856 Riemann used his theory to give a very elegant solution of the Jacobi inversion problem. Since an analytic function must be constant if it has no poles on a Riemann surface, it was possible to use the periods of the integrals that occur in the problem to determine the function up to a constant multiple and then to find quotients of theta functions having the same periods, thereby solving the problem.

41.2.6. Weierstrass

Of the three founders of analytic function theory, Weierstrass was the most methodical. He had found his own solution to the Jacobi inversion problem and submitted it simultaneously with Riemann. When he saw Riemann's work, he withdrew his own paper and spent many years working out in detail how the two approaches related to each other. Where Riemann had allowed his geometric intuition to build castles in the air, so to speak, Weierstrass was determined to anchor his ideas in a firm algebraic foundation. Instead of picturing kinematically a point wandering from one sheet of a Riemann surface to another, Weierstrass

preferred a static object that he called a *Gebilde* (*structure*). His *Gebilde* was based on the set of pairs of complex numbers (z, w) satisfying a polynomial equation $p(z, w) = 0$, where $p(z, w)$ was an irreducible polynomial in the two variables. These pairs were supplemented by certain ideal points of the form (z, ∞) , (∞, w) , or (∞, ∞) when one or both of w or z tended to infinity as the other approached a finite or infinite value. Around all but a finite set of points, it was possible to expand w in an ordinary Taylor series in nonnegative integer powers of $z - z_0$. For each of the exceptional points, there would be one or more expansions in fractional or negative powers of $z - z_0$, as Puiseux and Laurent had found. These power series were Weierstrass' basic tool in analytic function theory.

41.3. COMPARISON OF THE THREE APPROACHES

Cauchy's approach seems to subsume the work of both Riemann and Weierstrass. Riemann, to be sure, had a more elegant way of overcoming the difficulty presented by multivalued functions, but Cauchy and Puiseux between them came very close to doing something logically equivalent. Weierstrass began with the power series and considered only functions that have a power-series development. That requirement appears to eliminate a large number of functions from consideration at the very outset, whereas Cauchy assumed only that the function is continuously differentiable and only later showed that in fact it must have a power-series development.⁹

On the other hand, when this theory is *applied* to study a particular function, the apparently greater generality of Cauchy's approach seems less obvious. Before you can prove anything about the function using Cauchy's theorems, you must verify that the function is differentiable. In order to do that, you have to know the definition of the function. How is that definition to be communicated, if not through some formula like a power series or other well-known function whose analyticity is known? Weierstrass saw this point clearly; in 1884 he said, "No matter how you twist and turn, you cannot avoid using some sort of analytic expressions such as power series" (quoted by Siegmund-Schultze, 1988, p. 253).

PROBLEMS AND QUESTIONS

Mathematical Problems

41.1. The formula $\cos \theta = 4 \cos^3(\theta/3) - 3 \cos(\theta/3)$, can be rewritten as the equation $p(\cos \theta/3, \cos \theta) = 0$, where $p(x, y) = 4x^3 - 3x - y$. Observe that $\cos(\theta + 2m\pi) = \cos \theta$ for all integers m , so that

$$p\left(\cos\left(\frac{\theta + 2m\pi}{3}\right), \cos \theta\right) \equiv 0,$$

for all integers m . That makes it very easy to construct the roots of the equation $p(x, \cos \theta) = 0$. They must be $\cos((\theta + 2m\pi)/3)$ for $m = 0, 1, 2$. What is the analogous equation for dividing a circular arc into five equal pieces?

⁹Cauchy *assumed* that the derivative was continuous. It was later shown by Edouard Goursat (1858–1936) in 1900 that differentiability implies continuous differentiability on open subsets of the plane.

- 41.2.** Suppose (as is the case for elliptic functions) that $f(x)$ is doubly periodic, that is, $f(x + m\omega_1 + n\omega_2) = f(x)$ for all m and n . Suppose also that there is a polynomial $p(x)$ of degree n^2 such that $p(f(\theta/n)) = f(\theta)$ for all θ . Finally, suppose you know a number say φ , such that $f(\varphi) = C$. Show that the roots of the equation $p(x) = C$ must be $x_{m,n} = f(\varphi/n + (k/n)\omega_1 + (l/n)\omega_2)$, where k and l range independently from 0 to $n - 1$.
- 41.3.** Use L'Hospital's rule to verify that

$$\lim_{h \rightarrow 0} \frac{f(x + 2h) + f(x) - 2f(x + h)}{h^2} = f''(x)$$

for any function $f(x)$ such that $f''(x)$ is continuous. Then suppose that $f(x) = \sum a_n x^n$ and $g(x) = \sum b_n x^n$ are two analytic functions on the interval $-1 < x < 1$ such that $f(x) = g(x)$ for some interval $-\varepsilon < x < \varepsilon$ for some positive number ε , no matter how small. Prove that $f(0) = g(0)$, $f'(0) = g'(0)$, and $f''(0) = g''(0)$. (Similarly, it can be shown by using higher-order difference quotients of the function f to get its derivatives at 0 that $f^{(n)}(0) = g^{(n)}(0)$ for all n . It then follows that $a_n = f^{(n)}(0)/n! = g^{(n)}(0)/n! = b_n$, and therefore that $f(x) = g(x)$ everywhere. Thus, two analytic functions that coincide at all points in some interval around 0 must coincide everywhere.)

Historical Questions

- 41.4.** What mathematical problems forced mathematicians to take complex numbers seriously instead of rejecting them as unusable?
- 41.5.** What role did algebraic integrals play in the development of modern complex analysis?
- 41.6.** What differences exist among the approaches of Cauchy, Riemann, and Weierstrass to the theory of analytic functions of a complex variable?

Questions for Reflection

- 41.7.** We noted at the beginning of this chapter that convergent power series can represent only the ultra-smooth functions we call analytic, while convergent trigonometric series can represent functions that have very arbitrary breaks and kinks in their graphs. Considering that every complex number z has a polar representation $z = r(\cos \theta + i \sin \theta)$, and $z^n = r^n (\cos(n\theta) + i \sin(n\theta))$, how do you account for this difference?
- 41.8.** What value is there in starting from Cauchy's assumption that a function has a complex derivative at every point of a domain and then using the Cauchy integral to prove that it must then have a convergent power series expansion about each point? Weierstrass saw none, and preferred merely to start with the convergent power series as the *definition* of the function. Of course, it is easy to show that a convergent power series has a complex derivative at each point, so that the two definitions are equivalent. Is there an aesthetic or psychological reason for preferring one to the other?

- 41.9.** Lagrange championed the use of analytic functions in physics because of the property that all the derivatives of an analytic function at a point can be computed (using higher-order differences) from the values of the function in an arbitrarily small neighborhood of the point. If the independent variable is time, this implies that the values the function over any time interval, no matter how short, determine its value at all subsequent times. In short, analytic functions are deterministic. How does this property mesh with the assumptions of classical physics? How would today's physicists look at it?

Real Numbers, Series, and Integrals

In complex analysis attention is restricted from the outset to functions that have a complex derivative. That very strong assumption automatically ensures that the functions studied will have convergent Taylor series. If only mathematical physics could manage with just such smooth functions, the abstruse concepts that fill up courses in real analysis would not be needed. But the physical world is full of boundaries, where the density of matter is discontinuous, temperatures undergo abrupt changes, light rays reflect and refract, and vibrating membranes are clamped. For these situations the imaginary part of the variable, which often has no physical interpretation anyway, might as well be dropped, since its only mathematical role was to complete the analytic function. From that point on, analysis proceeds on the basis of real variables only. Real analysis, which represents another extension of calculus, has to deal with very general, “rough” functions. All of the logical difficulties about calculus poured into this area of analysis, including the important questions of convergence of series, existence of maxima and minima, allowable ways of defining functions, continuity, and the meaning of integration. As a result, real analysis is so much less unified than complex analysis that it hardly appears to be a single subject. Its basic theorems do not follow from one another in any canonical order, and their proofs tend to be a bag of special tricks, rarely remembered for long except by professors who lecture on the subject constantly.

The subject arose in the attempts to solve the partial differential equations of mathematical physics, the wave equation, the Laplace equation, and, later on, the heat equation. Thus, to speak paradoxically, its roots are in the branches of the subject. At the same time, real analysis techniques forced mathematicians to confront the issues of what is meant by an integral, in what sense a series converges, and what, in the final analysis, a real number actually *is*. Thus, the branches of real analysis extended into the roots of analysis in general.

The free range of intuition suffered only minor checks in complex analysis. In that subject, what one wanted to believe very often turned out to be true. But real analysis almost seemed to be trapped in a hall of mirrors at times, as it struggled to gain the freedom to operate while avoiding paradoxes and contradictions. The generality of operations allowed in real analysis has fluctuated considerably over the centuries. While Descartes had imposed rather strict criteria for allowable curves (functions), Daniel Bernoulli attempted to represent very arbitrary functions as trigonometric series, and the mathematical physicist André-Marie Ampère (1775–1836) attempted to prove that a continuous function (in the modern sense,

but influenced by preconceptions based on the earlier sense) would have a derivative at most points. The critique of this proof was followed by several decades of backtracking, as more and more exceptions were found for operations with series and integrals that appeared to be formally all right. Eventually, when a level of rigor was reached that eradicated the known paradoxes, the time came to reach for more generality. Georg Cantor's set theory played a large role in this increasing generality, while developing paradoxes of its own. In the twentieth century, the theories of generalized functions and distributions restored some of the earlier freedom by inventing a new object to represent the derivative of functions that have no derivative in the ordinary sense.

42.1. FOURIER SERIES, FUNCTIONS, AND INTEGRALS

There is a symmetry in the development of real and complex analysis. Broadly speaking, both arose from differential equations, and complex analysis grew out of power series, while real analysis grew out of trigonometric series. These two techniques, closely connected with each other through the relation $z^n = r^n(\cos n\theta + i \sin n\theta)$, led down divergent paths that nevertheless crossed frequently in their meanderings. The real and complex viewpoints in analysis began to diverge with the study of the vibrating string problem in the 1740s by d'Alembert, Euler, and Daniel Bernoulli.

For a string fastened at two points, say $(0, 0)$ and $(L, 0)$ and vibrating so that its displacement above or below the point $(x, 0)$ at time t is $y(x, t)$, mathematicians agreed that the best compromise between realism and comprehensibility to describe this motion was the *one-dimensional wave equation*, which d'Alembert studied in 1747,¹ publishing the results in 1749:

$$\frac{\partial^2 y}{\partial t^2} = c^2 \frac{\partial^2 y}{\partial x^2}.$$

D'Alembert exhibited a very general solution of this problem in the form

$$y(x, t) = \Psi(t + x) + \Gamma(t - x),$$

where for simplicity he assumed that $c = 1$. The equation alone does not determine the function, of course, since the vibrations depend on the initial position and velocity of the string.

The following year, Euler took up this problem and commented on d'Alembert's solution. He observed that the initial position could be any shape at all, "either regular or irregular and mechanical." D'Alembert found that claim hard to accept. After all, the functions Ψ and Γ had to have periodicity and parity properties. How else could they be defined except as power series containing only odd or only even powers? Euler and d'Alembert were not interpreting the word "function" in the same way. Euler was even willing to consider initial positions $f(x)$ with corners (a "plucked" string), whereas d'Alembert insisted that $f(x)$ must have two derivatives in order to satisfy the equation.

¹Thirty years earlier, Brook Taylor (1685–1731) had analyzed the problem geometrically and concluded that the normal acceleration at each point would be proportional to the normal curvature at that point. That statement is effectively the same as this equation, and it was quoted by d'Alembert.

Three years later, Daniel Bernoulli tried to straighten this matter out, giving a solution in the form

$$y(x, t) = \sum_{n=1}^{\infty} a_n \sin\left(\frac{n\pi x}{L}\right) \cos\left(\frac{n\pi ct}{L}\right),$$

which he did not actually write out. Here the coefficients a_n were to be chosen so that the initial condition was satisfied, that is,

$$f(x) = y(x, 0) = \sum_{n=1}^{\infty} a_n \sin\left(\frac{n\pi x}{L}\right).$$

Observing that he had an infinite set of coefficients at his disposal for “fitting” the function, Bernoulli claimed that “any” function $f(x)$ had such a representation. Bernoulli’s solution was the first of many instances in which the classical partial differential equations of mathematical physics—the wave, heat, and potential equations—were studied by separating variables and superposing the resulting solutions. The technique was ultimately to lead to what are called Sturm–Liouville problems.

Before leaving the wave equation, we must mention one more important intersection between real and complex analysis in connection with it. In studying the action of gravity, Pierre-Simon Laplace (1749–1827) was led to what is now known as Laplace’s equation in three variables. The two-variable version of this equation in rectangular coordinates—Laplace was using polar coordinates—is

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

The operator on the left-hand side of this equation is known as the *Laplacian*. Since Laplace’s equation can be thought of as the wave equation with velocity $c = \sqrt{-1}$, complex numbers again enter into a physical problem. Recalling d’Alembert’s solution of the wave equation, Laplace suggested that the solutions of his equation might be sought in the form $f(x + y\sqrt{-1}) + g(x - y\sqrt{-1})$. Once again a problem that started out as a real-variable problem led naturally to the need to study functions of a complex variable.

42.1.1. The Definition of a Function

Daniel Bernoulli accepted his father’s definition of a function as “an expression formed in some manner from variables and constants,” as did Euler and d’Alembert. But those words seemed to have different meanings for each of them. Daniel Bernoulli thought that his solution met the criterion of being “an expression formed from variables and constants.” His former colleague in the Russian Academy of Sciences,² Euler, saw the matter differently. This time it was Euler who argued that the concept of function was being used too loosely. According to him, since the right-hand side of Bernoulli’s formula consisted of odd functions of period $2L$, it could represent only an odd function of period $2L$. Therefore, he said, it

²Bernoulli had left St. Petersburg in 1733, Euler in 1741.

did not have the generality of the solution he and d'Alembert had given. Bottazzini (1986, p. 29) describes the situation concisely:

We are here facing a misunderstanding that reveals one aspect of the contradictions between the old and new theory of functions, even though they are both present in the same man, Euler, the protagonist of this transformation.

The difference between the old and new concepts is seen in the simplest example, the function $|x|$, which equals x when $x \geq 0$ and $-x$ for $x \leq 0$. We have no difficulty thinking of this function as one function. It appeared otherwise to nineteenth-century mathematicians. Fourier described what he called a “discontinuous function represented by a definite integral” in 1822: the function

$$\frac{2}{\pi} \int_0^{\infty} \frac{\cos qx}{1+q^2} dq = \begin{cases} e^{-x} & \text{if } x \geq 0, \\ e^x & \text{if } x \leq 0. \end{cases}$$

Fifty years later, Gaston Darboux (1844–1918) gave the modern point of view, that this function is not truly discontinuous but merely a function expressed by two different analytic expressions in different parts of its domain.

The change in point of view came about gradually, but an important step was Cauchy's refinement of the definition in the first chapter of his 1821 *Cours d'analyse*:

When variable quantities are related so that, given the value of one of them, one can infer those of the others, we normally consider that the quantities are all expressed in terms of one of them, which is called the *independent* variable, while the others are called *dependent variables*.

Cauchy's definition still does not specify what *ways* of expressing one variable in terms of another are legitimate, but this definition was a step toward the basic idea that the value of the independent variable determines (uniquely) the value of the dependent variable or variables.

42.2. FOURIER SERIES

Daniel Bernoulli's work introduced trigonometric series as an alternative to power series. In a classic work of 1811, a revised version of which was published in 1821,³ *Théorie analytique de chaleur* (*Analytic Theory of Heat*), Fourier established the standard formulas for the Fourier coefficients of a function. For an even function of period 2π , these formulas are

$$f(x) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n \cos nx, \quad a_n = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos nx \, dx, \quad n = 0, 1, \dots$$

³The original version remained unpublished until 1972, when Grattan-Guinness published an annotated version of it.

A trigonometric series whose coefficients are obtained from an integrable function $f(x)$ in this way is called a *Fourier series*.

42.2.1. Sturm–Liouville Problems

After trigonometric series had become a familiar technique, mathematicians were encouraged to look for other simple functions in terms of which solutions of more general differential equations than Laplace's equation could be expressed. Between 1836 and 1838 this problem was attacked by Charles Sturm (1803–1855) and Joseph Liouville, who considered general second-order differential equations of the form

$$[p(x)y'(x)]' + [\lambda r(x) + q(x)]y(x) = 0.$$

When a solution of Laplace's equation is sought in the form of a product of functions of one variable (the separation of variables technique), the result is an equation of this type for the one-variable functions. It often happens that only isolated values of λ yield solutions satisfying given boundary conditions. Sturm and Liouville found that in general there will be an infinite set of values $\lambda = \lambda_n$, $n = 1, 2, \dots$, satisfying the equation together with a pair of conditions at the endpoints of an interval $[a, b]$, and that these values increase to infinity. The values can be arranged so that the corresponding solutions $y_n(x)$ have exactly n zeros in $[a, b]$, and any solution of the differential equation can be expressed as a series

$$y(x) = \sum_{n=1}^{\infty} c_n y_n(x).$$

The sense in which such series converge was still not clear, but it continued to be studied by other mathematicians. It required some decades for all these ideas to be sorted out.

Proving that a Fourier series actually did converge to the function that generated it was one of the first places where real analysis encountered greater difficulties than complex analysis. In 1829 Peter Lejeune Dirichlet (1805–1859) proved that the Fourier series of $f(x)$ converged to $f(x)$ for a bounded periodic function $f(x)$ having only a finite number of discontinuities and a finite number of maxima and minima in each period.⁴ Dirichlet tried to get necessary and sufficient conditions for convergence, but that is a problem that has never been solved. He showed that some kind of continuity would be required by giving the famous example of the function whose value at x is one of two different values according as x is rational or irrational. This function is called the *Dirichlet function*. For such a function, he thought, no integral could be defined, and therefore no Fourier series could be defined.⁵

⁴We would call such a function *piecewise monotonic*.

⁵The increasing latitude allowed in analysis, mentioned above, is illustrated very well by this example. When the Lebesgue integral is used, this function is regarded as identical with the constant value it assumes on the irrational numbers.

42.3. FOURIER INTEGRALS

The convergence of the Fourier series of $f(x)$ can be expressed as the equation

$$f(x) = \frac{1}{\pi} \int_0^\pi f(y) dy + \frac{2}{\pi} \sum_{n=1}^{\infty} \int_0^\pi f(y) \cos(ny) \cos(nx) dy.$$

That equation may have led to an analogous formula for Fourier integrals, which appeared during the early nineteenth century in papers on the wave and heat equations written by Poisson, Laplace, Fourier, and Cauchy. The central discovery in this area was the Fourier inversion formula, which we now write as

$$f(x) = \frac{2}{\pi} \int_0^\infty \int_0^\infty f(y) \cos(zy) \cos(zx) dy dz.$$

The analogy with the formula for series is clear: The continuous variable z replaces the discrete index n , and the integral on z replaces the sum over n . Once again, the validity of the representation is much more questionable than the validity of the formulas of complex analysis, such as the Cauchy integral formula for an analytic function. The Fourier inversion formula has to be interpreted very carefully, since the order of integration cannot be reversed. If the integrals make sense in the order indicated, that happy outcome can only be the result of some special properties of the function $f(x)$. But what are those properties?

The difficulty was that the integral extended over an infinite interval so that convergence required the function to have two properties: It needed to be continuous, and it needed to decrease sufficiently rapidly at infinity to make the integral converge. These properties turned out to be, in a sense, dual to each other. Considering just the inner integral as a function of z :

$$\widehat{f}(z) = \int_0^\infty f(y) \cos(zy) dy,$$

it turns out that the more rapidly $f(y)$ decreases at infinity, the more derivatives $\widehat{f}(z)$ has, and the more derivatives $f(y)$ has, the more rapidly $\widehat{f}(z)$ decreases at infinity. The converses are also, broadly speaking, true. Could one insist on having both conditions, so that the representation would be valid? Would these assumptions impair the usefulness of these techniques in mathematical physics? Alfred Pringsheim (1850–1941, father-in-law of the writer Thomas Mann) studied the Fourier integral formula (Pringsheim, 1910), noting especially the two kinds of conditions that $f(x)$ needed to satisfy, which he called “conditions in the finite region” (“im Endlichen”) and “conditions at infinity” (“im Unendlichen”). Nowadays, they are called local and global conditions. Pringsheim noted that the local conditions could be traced all the way back to Dirichlet’s work of 1829, but said that “a rather obvious backwardness reveals itself” in regard to the global conditions.

[They] seem in general to be limited to a relatively narrow condition, one which is insufficient for even the simplest type of application, namely that of absolute integrability of $f(x)$ over an infinite interval. There are, as far as I know, only a few exceptions.

Thus, to the question as to whether physics could get by with sufficiently smooth functions $f(x)$ that decay sufficiently rapidly, the answer turned out to be, in general, no. Physics needs to deal with discontinuous integrable functions $f(y)$, and for these $f(z)$ cannot decay rapidly enough at infinity to make its integral converge absolutely. What was to be done?

One solution involved the introduction of *convergence factors*, leading to a more general sense of convergence, called Abel–Poisson convergence. In a paper on wave motion published in 1818, Siméon-Denis Poisson (1780–1840) used the representation

$$f(x) = \frac{1}{\pi} \int_0^\infty \int_{-\infty}^{+\infty} f(\alpha) \cos a(x - \alpha) e^{-ka} d\alpha da.$$

The exponential factor provided enough decrease at infinity to make the integral converge. Poisson claimed that the resulting integral tended toward $f(x)$ as k decreased to 0. (He was right.)

Abel used an analogous technique with infinite series, multiplying the n th term by r^n , where $0 < r < 1$, then letting r increase to 1. In this way, he was able to justify the natural value assigned to some nonabsolutely convergent series such as

$$\ln(2) = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots \quad \text{and} \quad \frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \dots,$$

which can be obtained by expanding the integrands of the following integrals as geometric series and integrating termwise:

$$\int_0^1 \frac{1}{1+r} dr, \quad \int_0^1 \frac{1}{1+r^2} dr.$$

In Abel's case, the motive for making a careful study of continuity was his having noticed that a trigonometric series could represent a discontinuous function. From Paris in 1826 he wrote to a friend that the expansion

$$\frac{x}{2} = \sin x - \frac{1}{2} \sin 2x + \frac{1}{3} \sin 3x - \frac{1}{4} \sin 4x + \dots$$

was provable for $0 \leq x < \pi$, although obviously it could not hold at $x = \pi$. Thus, while the representation might be a good thing, it meant, on the other hand, that the sum of a series of continuous functions could be discontinuous. Abel also believed that many of the difficulties mathematicians were encountering were traceable to the use of divergent series. He gave, accordingly, a thorough discussion of the convergence of the binomial series, the most difficult of the elementary Taylor series to analyze.⁶

For the two conditionally convergent series shown above and the general Fourier integral, continuity of the sum was needed. In both cases, what appeared to be a necessary evil—the introduction of the convergence factor e^{-ka} or r^n —turned out to have positive value.

⁶Unknown to Abel, Bolzano had discussed the binomial series in 1816, considering integer, rational, and irrational (real) exponents, admitting that he could not cover all possible cases, due to the incomplete state of the theory of complex numbers at the time (Bottazzini, 1986, pp. 96–97). He performed a further analysis of series in general in 1817, with a view to proving the intermediate value property).

For the functions $r^n \cos n\theta$ and $r^n \sin n\theta$ are harmonic functions if r and θ are regarded as polar coordinates, while $e^{-ay} \cos(ax)$ and $e^{-ay} \sin(ax)$ are harmonic if x and y are regarded as rectangular coordinates. The factors used to ensure convergence provided harmonic functions, at no extra cost.

42.4. GENERAL TRIGONOMETRIC SERIES

The study of trigonometric functions advanced real analysis once again in 1854, when Riemann was required to give a lecture to qualify for the position of *Privatdocent* (roughly what would be an assistant professor nowadays). As the rules required, he was to propose three topics and the faculty would choose the one he lectured on. One of the three, based on conversations he had had with Dirichlet over the preceding year, was the representation of functions by trigonometric series.⁷ Dirichlet was no doubt hoping for more progress toward necessary and sufficient conditions for convergence of a Fourier series, the topic he had begun so promisingly a quarter-century earlier. Riemann concentrated on one question in particular: *Can a function be represented by more than one trigonometric series?* That is, *can two trigonometric series with different coefficients have the same sum at every point?* The importance of this problem seems to come from the possibility of starting with a general trigonometric series and summing it. One then has a periodic function which, if it is sufficiently smooth, is the sum of its Fourier series. The natural question arises: Is that Fourier series the trigonometric series that generated the function in the first place?

In the course of his study, Riemann was driven to examine the fundamental concept of integration. Cauchy had defined the integral

$$\int_a^b f(x) dx$$

as the number approximated by the sums

$$\sum_{n=1}^N f(x_n)(x_n - x_{n-1})$$

as N becomes large, where $a = x_0 < x_1 < \cdots < x_{N-1} < x_N = b$. Riemann refined the definition slightly, allowing $f(x_n)$ to be replaced by $f(x_n^*)$ for any x_n^* between x_{n-1} and x_n . The resulting integral is known as the Riemann integral today. Riemann sought necessary and sufficient conditions for such an integral to exist. The condition that he formulated led ultimately to the concept of a set of measure zero,⁸ half a century later: *For each $\sigma > 0$ the total length of the intervals on which the function $f(x)$ oscillates by more than σ must become arbitrarily small if the partition is sufficiently fine.*

⁷As the reader will recall from Chapter 40, this topic was *not* the one Riemann did lecture on. Gauss preferred the topic of foundations of geometry, and so Riemann's paper on trigonometric series was not published until 1867, after his death.

⁸A set of points on the line has measure zero if for every $\varepsilon > 0$ it can be covered by a sequence of intervals (a_k, b_k) whose total length is less than ε .

PROBLEMS AND QUESTIONS

Mathematical Problems

- 42.1.** Show that if $y(x, t) = (f(x + ct) + f(x - ct))/2$ is a solution of the one-dimensional wave equation that is valid for all x and t , and $y(0, t) = 0 = y(L, t)$ for all t , then $f(x)$ must be an odd function of period $2L$.
- 42.2.** Show that the problem $X''(x) - \lambda X(x) = 0$, $Y''(y) + \lambda Y(y) = 0$, with boundary conditions $Y(0) = Y(2\pi)$, $Y'(0) = Y'(2\pi)$, implies that $\lambda = n^2$, where n is an integer, and that the function $X(x)Y(y)$ must be of the form $(c_n e^{nx} + d_n e^{-nx})(a_n \cos(ny) + b_n \sin(ny))$ if $n \neq 0$.
- 42.3.** Show that Fourier series can be obtained as the solutions to a Sturm–Liouville problem on $[0, 2\pi]$ with $p(x) = r(x) \equiv 1$, $q(x) = 0$, with the boundary conditions $y(0) = y(2\pi)$, $y'(0) = y'(2\pi)$. What are the possible values of λ ?

Historical Questions

- 42.4.** Why did the problem of the vibrating string force the consideration of nonanalytic solutions of differential equations?
- 42.5.** What problems arose in the use of trigonometric series that had not arisen in the use of power series?
- 42.6.** How did Sturm–Liouville problems come to be an area of particular interest in analysis?

Questions for Reflection

- 42.7.** What is the value of harmonic functions, which are solutions of Laplace’s equation

$$0 = \nabla^2 u(x) = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}.$$

Consider that the classical linearized heat equation describes the temperature $u(t; x, y)$ at point (x, y) of a plate at time t is

$$\frac{\partial u}{\partial t} = k \nabla^2 u(x) = k \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right),$$

and that the classical linearized wave equation describes the vertical displacement of a membrane $u(t; x, y)$ over a point (x, y) of a plate at time t is

$$\frac{\partial^2 u}{\partial t^2} = c^2 \nabla^2 u(x) = c^2 \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right),$$

where c is the velocity of wave propagation in the membrane. (What does it mean for the first or second derivative with respect to time to be zero?)

- 42.8.** The Cauchy–Riemann equations (see Chapter 41) and the equality of mixed partial derivatives ($\partial^2 u / \partial x \partial y = \partial^2 u / \partial y \partial x$) easily imply that the real and imaginary parts of an analytic function of a complex variable are harmonic functions. Does it follow that if $u(x, y)$ and $v(x, y)$ are real-valued harmonic functions, then $u(x, y) + iv(x, y)$ is an analytic function? What further property is needed?
- 42.9.** Why is it of interest to know whether two different trigonometric series can converge to the same function?

Foundations of Real Analysis

The uncritical use of limiting processes, which did little harm when applied to analytic functions of a complex variable, led to acute problems in the case of general functions of a real variable. The attempts to avoid self-contradictory results led to a close scrutiny of the properties of the real line and the identification of certain hidden assumptions that were needed to establish standard results, such as Cauchy's "theorem" that the sum of an infinite series of continuous functions is continuous. This close scrutiny, in turn, led to set theory. Trigonometric series were involved at the beginning of set theory, although nowadays it is developed without any reference to them. We shall discuss set theory in the next chapter. The present chapter is devoted to that closer scrutiny of the real numbers.

43.1. WHAT IS A REAL NUMBER?

We have already alluded to the fact that our modern notion of a real number as an infinite decimal expansion really goes back to the Eudoxan concept of a ratio. A great advance came in the seventeenth century, when analytic geometry was invented by Descartes and Fermat. In his *Géométrie*, Descartes showed how to replace a ratio *in thought* by a line, choosing a line arbitrarily called a unit, and letting any other line stand for the number represented by its ratio to that unit.

Euclid had not discussed the product of two lines. He spoke instead of the rectangle on the two lines. Stimulated by algebra, however, and the application of geometry to it, Descartes looked at the product of two lengths in a different way. As pure numbers, the product ab is simply the fourth proportional to $1 : a : b$. That is, $ab : b :: a : 1$. He therefore fixed an arbitrary line that he called I to represent the number 1 and represented ab as the line that satisfied the proportion $ab : b :: a : I$, when a and b were lines representing two given numbers.

The notion of a *real number* had at last arisen, not as most people think of it today—an infinite decimal expansion—but as a ratio of line segments. Only a few decades later Newton defined a real number to be "the ratio of one magnitude to another magnitude of the same kind, arbitrarily taken as a unit." Newton classified numbers as integers, fractions, and surds (Whiteside, 1967, Vol. 2, p. 7). Even with this clarification, however, mathematicians were inclined to gloss over certain difficulties. For example, there is an arithmetic rule according to which $\sqrt{ab} = \sqrt{a}\sqrt{b}$. In Descartes' geometric definition of the product of two real

numbers, it is not obvious how this rule is to be proved. The use of the decimal system, with its easy approximations to irrational numbers, soothed the consciences of mathematicians and gave them the confidence to proceed with their development of the calculus. No one even seemed very concerned about the absence of any good geometric construction of cube roots and higher roots of real numbers. The real line answered the needs of algebra in that it gave a representation of any real root there might be of any algebraic equation with real numbers as coefficients. It was some time before anyone realized that geometry still had resources that even algebra did not encompass and would lead to numbers for which pure algebra had no use.

Those resources included the continuity of the geometric line, which turned out to be exactly what was needed for the limiting processes of calculus. It was this property that made it sensible for Euler to talk about the number that we now call e , that is,

$$e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = \sum_{n=0}^{\infty} \frac{1}{n!} = 2.7182818284590\dots,$$

and the other Euler constant

$$\gamma = \lim_{n \rightarrow \infty} \left[\left(\sum_{k=1}^n \frac{1}{k} \right) - \log n \right] = 0.5772156649\dots$$

The intuitive notion of continuity assured mathematicians that there were points on the line, and hence infinite decimal expansions, that must represent these numbers, even though no one would ever know the full expansions. The geometry of the line provided a geometric representation of real numbers and made it possible to reason about them without having to worry about the decimal expansion.

The continuity of the line brought the realization that the real numbers had more to offer than merely convenient representations of the solutions of equations. They could even represent some numbers such as e and γ that had not been found to be solutions of any equations. The line is richer than it needs to be for algebra alone. The concept of a real number allows arithmetic to penetrate into parts of geometry where even algebra cannot go. The sides and diagonals of regular figures such as squares, cubes, pentagons, pyramids, and the like all have ratios that can be represented as the solutions of equations, and hence are algebraic. For example, the diagonal D and side S of a pentagon satisfy $D^2 = S(D + S)$. For a square the relationship is $D^2 = 2S^2$, and for a cube it is $D^2 = 3S^2$. But what about the number we now call π , the ratio of the circumference C of a circle to its diameter D ? In the seventeenth century, Leibniz noted that any line that could be constructed using Euclidean methods (straightedge and compass) would have a length that satisfied some equation with rational coefficients. In a number of letters and papers written during the 1670s, Leibniz was the first to contrast what is algebraic (involving polynomials with rational coefficients) with objects that he called *analytic* or *transcendental* and the first to suggest that π might be transcendental. In the preface to his pamphlet *De quadratura arithmetica circuli* (*On the Arithmetical Quadrature of the Circle*), he explained that at least as a function of a variable, the sine is *not* a polynomial:

A complete quadrature would be one that is both analytic and linear; that is, it would be constructed by the use of curves whose equations are of [finite] degrees. The illustrious [James] Gregory, in his book *On the Exact Quadrature of the Circle*, has claimed that this is impossible, but, unless I am mistaken, has given no proof. I still do not see what prevents the circumference itself, or some particular part of it, from being measured [that is, being commensurable with the radius], a part whose arc has a ratio to its sine [half-chord] that can be expressed by an equation of finite degree. But to express the ratio of the arc to the sine *in general* by an equation of finite degree is impossible, as I shall prove in this little work. [Gerhardt, 1971, Vol. 5, p. 97]

No representation of π as the root of a polynomial with rational coefficients was ever found. This ratio had a long history of numerical approximations from all over the world, but no one ever found any nonidentical equation with rational coefficients satisfied by the circumference and diameter of a circle. The fact that π is transcendental was first proved in 1881 by Ferdinand Lindemann (1852–1939). The complete set of real numbers thus consists of the positive and negative rational numbers, all real roots of equations with integer coefficients (the *algebraic* numbers), and the transcendental numbers. All transcendental numbers and some algebraic numbers are irrational. Examples of transcendental numbers turned out to be rather difficult to produce. The first well-known¹ number to be proved transcendental was the base of natural logarithms e , and this proof was achieved only in 1873, by the French mathematician Charles Hermite (1822–1901). It is still not known whether the Euler constant $\gamma \approx 0.57712$ is even irrational.

43.1.1. The Arithmetization of the Real Numbers

Not until the nineteenth century, when mathematicians took a retrospective look at the magnificent edifice of calculus that they had created and tried to give it the same degree of logical rigor possessed by algebra and Euclidean geometry, were attempts made to define real numbers arithmetically, without mentioning ratios of lines. One such definition by Richard Dedekind (1831–1916), a professor at the Zürich Polytechnikum, was inspired by a desire for rigor when he began lecturing to students in 1858. He found the rigor he sought without much difficulty, but did not bother to publish what he regarded as mere common sense until 1872, when he wished to publish something in honor of his father. In his book *Stetigkeit und irrationale Zahlen (Continuity and Irrational Numbers)* he referred to Newton's definition of a real number:

... the way in which the irrational numbers are usually introduced is based directly upon the conception of extensive magnitudes—which itself is nowhere carefully defined—and explains number as the result of measuring such a magnitude by another of the same kind. Instead of this I demand that arithmetic shall be developed out of itself.

As Dedekind saw the matter, it was really the *totality* of rational numbers that defined a ratio of continuous magnitudes. Although one might not be able to say that two continuous quantities a and b had a ratio *equal* to, or defined by, a ratio $m : n$ of two integers, an inequality such as $ma < nb$ could be interpreted as saying that the real number $a : b$ (whatever it was) was *less than* the rational number n/m . In fact, that interpretation of the

¹Joseph Liouville had shown how to *construct* real numbers that are transcendental as early as 1844, using infinite series and continued fractions, but none of the numbers so constructed had ever arisen naturally.

inequality was the basis for the Eudoxan theory of proportion, although neither Eudoxus nor Euclid was able to say precisely what a ratio of two lines *is*.

Thus a positive real number could be defined as a way of dividing the positive rational numbers into two classes, those that were larger than the number and those that were equal to it or smaller, and every member of the first class was larger than every member of the second class. But, so reasoned Dedekind, once the positive rational numbers have been partitioned in this way, the *two classes themselves* can be regarded as the number.² They are a well-defined object, and one can define arithmetic operations on such classes so that the resulting system has all the properties we want the real numbers to have, especially the essential one for calculus: continuity. Dedekind claimed that in this way he was able to prove rigorously for the first time that $\sqrt{2}\sqrt{3} = \sqrt{6}$.³

The practical-minded reader who is content to use approximations will probably be getting somewhat impatient with the discussion at this point and asking if it was really necessary to go to so much trouble to satisfy a pedantic desire for rigor. Such a reader will be in good company. Many prominent mathematicians of the time asked precisely that question. One of them was Rudolf Lipschitz (1832–1903). Lipschitz didn't see what the fuss was about, and he objected to Dedekind's claims of originality (Scharlau, 1986, p. 58). In 1876 he wrote to Dedekind:

I do not deny the validity of your definition, but I am nevertheless of the opinion that it differs only in form, not in substance, from what was done by the ancients. I can only say that I consider the definition given by Euclid... to be just as satisfactory as your definition. For that reason, I wish you would drop the claim that such propositions as $\sqrt{2}\sqrt{3} = \sqrt{6}$ have never been proved. I think the French readers especially will share my conviction that Euclid's book provided necessary and sufficient grounds for proving these things.

Dedekind refused to back down. He replied (Scharlau, 1986, pp. 64–65):

I have never imagined that my concept of the irrational numbers has any particular merit; otherwise I should not have kept it to myself for nearly fourteen years. Quite the reverse, I have always been convinced that any well-educated mathematician who seriously set himself the task of developing this subject rigorously would be bound to succeed... Do you really believe that such a proof can be found in any book? I have searched through a large collection of works from many countries on this point, and what does one find? Nothing but the crudest circular reasoning, to the effect that $\sqrt{a}\sqrt{b} = \sqrt{ab}$ because $(\sqrt{a}\sqrt{b})^2 = (\sqrt{a})^2(\sqrt{b})^2 = ab$; not the slightest explanation of how to multiply two irrational numbers. The proposition $(mn)^2 = m^2n^2$, which is proved for rational numbers, is used unthinkingly for irrational numbers. Is it not scandalous that the teaching of mathematics in schools is regarded as a particularly good means to develop the power of reasoning, while no other discipline (for example, grammar) would tolerate such gross offenses against logic for a minute? If one is to proceed scientifically, or cannot do so for lack of time, one should at least honestly tell the pupil to believe a proposition on the word of the teacher, which the students are willing to do anyway. That is better than destroying the pure, noble instinct for correct proofs by giving spurious ones.

²Actually, since the two classes determine each other, *one* of them, say the one consisting of larger numbers, can be taken as the definition of the real number. Thus $\sqrt{2}$ can be defined as the set of all positive rational numbers r such that $r^2 > 2$.

³In his paper (1992) David Fowler (1937–2004) investigated a number of approaches to the arithmetization of the real numbers and showed how the specific equation $\sqrt{2}\sqrt{3} = \sqrt{6}$ could have been proved geometrically and also how difficult this proof would have been using many other natural approaches.

Mathematicians have accepted the need for Dedekind's rigor in the teaching of mathematics majors, although the idea of defining real numbers as partitions of the rational numbers (Dedekind cuts) is no longer the most popular approach to that rigor. More often, students are now given a set of axioms for the real numbers and asked to accept on faith that those axioms are consistent and that they characterize a set that has the properties of a geometric line. Only a few books attempt to start with the rational numbers and construct the real numbers. Those that do tend to follow an alternative approach, defining a real number to be a sequence of rational numbers (more precisely, an equivalence class of such sequences, one of which is the sequence of successive decimal approximations to the number).

43.2. COMPLETENESS OF THE REAL NUMBERS

Dedekind's arithmetization of the real numbers amounted to the statement that the real numbers form a complete metric space. The concept now known as completeness of the real numbers is associated with the *Cauchy convergence criterion*, which asserts that a sequence of real numbers $\{a_n\}_{n=1}^{\infty}$ converges to some real number a if it is a *Cauchy sequence*; that is, for every $\varepsilon > 0$ there is an index n such that $|a_n - a_k| < \varepsilon$ for all $k \geq n$. This condition was stated somewhat loosely by Cauchy in his *Cours d'analyse*, published in the mid-1820s, and the proof given there was also somewhat loose. The same criterion had been stated, and for sequences of functions rather than sequences of numbers, a decade earlier by Bernhard Bolzano (1741–1848). In imprecise language, this criterion says that there exists a number that the sequence is getting close to, provided its terms are getting close to one another. The point is that the criterion of getting close to one another makes no reference to anything outside the sequence. Without this criterion, it would presumably be necessary to exhibit the limit explicitly in order to prove that the sequence converges. That might be difficult to do. Indeed, it *is* difficult in the case of such Dirichlet series as

$$\sum_{n=1}^{\infty} \frac{1}{n^3},$$

whose partial sums get close to one another, but whose sum has never been expressed in finite terms using only known real numbers.

43.3. UNIFORM CONVERGENCE AND CONTINUITY

Cauchy was not aware at first of any need to make the distinction between pointwise and uniform convergence, and he even claimed that the sum of a series of continuous functions would be continuous, a claim contradicted by Abel, as we have seen. The distinction is a subtle one. It is all too easy not to notice whether choosing n large enough to get a good approximation when $f_n(x)$ converges to $f(x)$ requires one to take account of which x is under consideration. That point needed to be stated precisely. The first clear statement of it is due to Philipp Ludwig von Seidel (1821–1896), a professor at Munich, who in 1847 studied the examples of Dirichlet and Abel, coming to the following conclusion:

When one begins from the certainty thus obtained that the proposition cannot be generally valid, then its proof must basically lie in some still hidden supposition. When this is subject to a precise analysis, then it is not difficult to discover the hidden hypothesis. One can then reason backwards that this [hypothesis] cannot occur [be fulfilled] with series that represent discontinuous functions. [Quoted in Bottazzini, 1986, p. 202]

In order to reason confidently about continuity, derivatives, and integrals, mathematicians began restricting themselves to cases where the series converged uniformly, that is, given a positive number ε , one could find an index N such that $|f_n(x) - f(x)| < \varepsilon$ for all $n > N$ and all x . Weierstrass, in particular, provided a famous theorem known as the M -test for uniform convergence of a series. But, although the M -test is certainly valuable in dealing with power series, uniform convergence in general is too severe a restriction. The trigonometric series exhibited by Abel, for example, represented a discontinuous function as the sum of a series of continuous functions and therefore did not converge uniformly. Yet it could be integrated term by term. One could provide many examples of series of continuous functions that converged to a continuous function but not uniformly. Weaker conditions were needed that would justify the operations rigorously without restricting their applicability too strongly.

43.4. GENERAL INTEGRALS AND DISCONTINUOUS FUNCTIONS

The search for less restrictive hypotheses and the consideration of more general figures on a line than just points and intervals led to more general notions of length, area, and integral, allowing more general functions to be integrated. Analysts began generalizing the integral beyond the refinements introduced by Riemann. Foundational problems also added urgency to this search. For example, in 1881, Vito Volterra (1860–1940) gave an example of a continuous function having a derivative at every point, but whose derivative was not Riemann integrable. What could the fundamental theorem of calculus mean for this derivative, which had an antiderivative but no integral, as integrals were then understood?

New integrals were created by the Latvian mathematician Axel Harnack (1851–1888), by the French mathematicians Emile Borel (1871–1956), Henri Lebesgue (1875–1941), and Arnaud Denjoy (1884–1974), and by the German mathematician Oskar Perron (1880–1975). By far the most influential of these was the Lebesgue integral, which was developed between 1899 and 1902. This integral was to have profound influence in the area of probability, due to its use by Borel, and in trigonometric series representations, an application that Lebesgue developed as an application of it. Lebesgue justified his more general integral in the preface to a 1904 monograph in which he expounded it, saying,

[I]f we wished to limit ourselves always to these good [that is, smooth] functions, we would have to give up on the solution of a number of easily stated problems that have been open for a long time. It was the solution of these problems, rather than a love of complications, that caused me to introduce in this book a definition of the integral that is more general than that of Riemann and contains the latter as a special case.

Despite its complexity—to develop it with proofs takes four or five times as long as developing the Riemann integral—the Lebesgue integral was included in textbooks as early as 1907: for example, *Theory of Functions of a Real Variable*, by E. W. Hobson (1856–1933). Its chief attraction was the greater generality of the conditions under which it allowed

termwise integration. For example, one of the main theorems of this theory is the Lebesgue dominated convergence theorem, which states that if $f_n(x)$ are measurable functions⁴ such that $|f_n(x)| \leq g(x)$, where $g(x)$ is integrable (in the sense of Lebesgue), and $f_n(x) \rightarrow f(x)$ pointwise, then $\int f_n(x) dx \rightarrow \int f(x) dx$. (Part of the theorem is that $f(x)$ is integrable. This is the Lebesgue dominated convergence theorem.)

Following the typical pattern of development in real analysis, the Lebesgue integral soon generated new questions. The Hungarian mathematician Frigyes Riesz (1880–1956) introduced the classes now known as L_p -spaces, the spaces of measurable functions f for which $|f|^p$ is Lebesgue integrable, $p > 0$. (The space L_∞ consists of functions that are bounded on a set whose complement has measure zero.) How the Fourier series and integrals of functions in these spaces behave became a matter of great interest, and a number of questions were raised. For example, in his 1915 dissertation at the University of Moscow, Nikolai Nikolaevich Luzin (1883–1950) posed the conjecture that the Fourier series of a (Lebesgue-) square-integrable function converges except on a set of measure zero. Fifty years elapsed before this conjecture was proved by the Swedish mathematician Lennart Carleson (b. 1928). Because a Riemann-integrable function is square-integrable in the sense of Lebesgue, this theorem applies in particular to such functions. Luzin's student Andrei Nikolaevich Kolmogorov (1903–1987) showed that the Fourier series of a function that is merely Lebesgue-integrable may diverge at every point.

43.5. THE ABSTRACT AND THE CONCRETE

The increasing generality allowed by the notation $y = f(x)$ threatened to carry mathematics off into stratospheric heights of abstraction. Although the mathematical physicist Ampère (1775–1836) had tried to show that a continuous function is differentiable at most points, the attempt was doomed to failure. Bolzano constructed a “sawtooth” function in 1817 that was continuous, yet had no derivative at any point. Weierstrass later used an absolutely convergent trigonometric series to achieve the same result,⁵ and a young Italian mathematician, Salvatore Pincherle (1853–1936), who took Weierstrass' course in 1877–1878, wrote a treatise in 1880 in which he gave a very simple example of such a function (Bottazzini, 1986, p. 286):

$$f(x) = \sum_{n=1}^{\infty} \frac{\sin(n!x)}{n!}.$$

Volterra's example of a continuous function whose derivative was not (Riemann) integrable, together with the examples of continuous functions having no derivative at any point naturally cast some doubt on the usefulness of the abstract concept of continuity and even the abstract concept of a function. Besides the construction of more general integrals and the consequent ability to “measure” more complicated geometric figures, it was necessary to investigate differentiation in more detail as well.

⁴See below for the definition of measurable functions.

⁵This example was communicated by his student Paul Du Bois-Reymond (1831–1889) in 1875. The following year Du Bois-Reymond constructed a continuous periodic function whose Fourier series failed to converge at a set of points that came arbitrarily close to every point.

43.5.1. Absolute Continuity

The secret of that quest turned out to be not continuity, but monotonicity. A continuous function may fail to have a derivative, but in order to fail, it must oscillate very wildly, as the examples of Bolzano and Weierstrass did. A function that did not oscillate, or oscillated only a finite total amount, necessarily had a derivative except on a set of measure zero. The ultimate result in this direction was achieved by Lebesgue, who showed that a monotonic function has a derivative on a set whose complement has measure zero. Such a function might or might not be the integral of its derivative, as the fundamental theorem of calculus states. In 1902 Lebesgue gave necessary and sufficient conditions for the fundamental theorem of calculus to hold; a function that satisfies these conditions, and is consequently the integral of its derivative, is called *absolutely continuous*.

43.5.2. Taming the Abstract

It had been known at least since the time of Lagrange that any finite set of n data points (x_k, y_k) , $k = 1, \dots, n$, with x_k all different, could be fitted perfectly with a polynomial of degree at most $n - 1$. Such a polynomial might—indeed, probably would—oscillate wildly in the intervals between the data points. Weierstrass showed in 1884 that any continuous function, no matter how abstract, could be uniformly approximated by a polynomial over any bounded interval $[a, b]$. Since there is always some observational error in any set of data, this result meant that polynomials could be used in both practical and theoretical ways, to fit data, and to establish general theorems about continuous functions. Weierstrass also proved a second version of the theorem, for periodic functions, in which he showed that for these functions the polynomial could be replaced by a finite sum of sines and cosines. This connection to the classical functions freed mathematicians to use the new abstract functions, confident that in applications they could be replaced by computable functions.

Weierstrass lived before the invention of the new abstract integrals mentioned above, although he did encourage the development of the abstract set theory of Georg Cantor, which provided the language in which these integrals were formulated. With the development of the Lebesgue integral, a new category of functions arose, the *measurable functions*. These are functions $f(x)$ such that the set of x for which $f(x) > c$ always has a meaningful measure, although it need not be a geometrically simple set, as it is in the case of continuous functions. It appeared that Weierstrass' work needed to be repeated, since his approximation theorem did not apply to measurable functions. In his 1915 dissertation, Luzin produced two beautiful theorems in this direction. The first was what is commonly called by his name nowadays, the theorem that for every measurable function $f(x)$ and every $\varepsilon > 0$ there is a continuous function $g(x)$ such that $g(x) \neq f(x)$ only on a set of measure less than ε . As a consequence of this result and Weierstrass' approximation theorem, it followed that every measurable function is the limit of a sequence of polynomials on a set whose complement has measure zero. Luzin's second theorem was that every finite-valued measurable function is the *derivative* of a continuous function at the points of a set whose complement has measure zero. He was able to use this result to show that any prescribed set of measurable boundary values on the disk could be the boundary values of a harmonic function.

With the Weierstrass approximation theorem and theorems like those of Luzin, modern analysis found some anchor in the concrete analysis of the "classical" period that ran from 1700 to 1850. But the striving for generality and freedom of operation still led to the invocation of some strong principles of inference in the context of set theory. By mid-twentieth

century mathematicians were accustomed to proving concrete facts using abstract techniques. To take just one example, it can be proved that some differential equations have a solution because a contraction mapping of a complete metric space must have a fixed point. Classical mathematicians would have found this proof difficult to accept, and many twentieth-century mathematicians have preferred to write in “constructivist” ways that avoid invoking the abstract “existence” of a mathematical object that cannot be displayed explicitly. But most mathematicians are now comfortable with such reasoning.

43.6. DISCONTINUITY AS A POSITIVE PROPERTY

The Weierstrass approximation theorems imply that the property of being the limit of a sequence of continuous functions is no more general than the property of being the limit of a sequence of polynomials or the sum of a trigonometric series. That fact raises an obvious question: What kind of function *is* the limit of a sequence of continuous functions? Du Bois-Reymond had shown that it can be discontinuous on a set that is, as we now say, dense. But can it, for example, be discontinuous at *every* point? That was one of the questions that interested René-Louis Baire (1874–1932). If one thinks of discontinuity as simply the absence of continuity, classifying mathematical functions as continuous or discontinuous seems to make no more sense than classifying mammals as cats or noncats. Baire, however, looked at the matter differently. In his 1905 *Leçons sur les fonctions discontinues (Lectures on Discontinuous Functions)* he wrote

Is it not the duty of the mathematician to begin by studying in the abstract the relations between these two concepts of continuity and discontinuity, which, while mutually opposite, are intimately connected?

Strange as this view may seem at first, we may come to have some sympathy for it if we think of the dichotomy between the continuous and the discrete, that is, between geometry and arithmetic. At any rate, to a large number of mathematicians at the turn of the twentieth century, it did not seem strange. The Moscow mathematician Nikolai Vasilevich Bugaev (1837–1903, father of the writer Andrei Belyi) was a philosophically inclined scholar who thought it possible to establish two parallel theories, one for continuous functions, the other for discontinuous functions. He called the latter theory *arithmology* to emphasize its arithmetic character. There is at least enough of a superficial parallel between integrals and infinite series and between continuous and discrete probability distributions (another area in which Russia has produced some of the world’s leaders) to make such a program plausible. It is partly Bugaev’s influence that caused works on set theory to be translated into Russian during the first decade of the twentieth century and brought the Moscow mathematicians Luzin and Dmitrii Fyodorovich Egorov (1869–1931) and their students to prominence in the area of measure theory, integration, and real analysis.

Baire’s monograph was single-mindedly dedicated to the pursuit of one goal: to give a necessary and sufficient condition for a function to be the pointwise limit of a sequence of continuous functions. He found the condition, building on earlier ideas introduced by Hermann Hankel (1839–1873): The necessary and sufficient condition is that the discontinuities of the function form a set of *first category*. A set is of first category if it is the union of a sequence of sets A_k such that every interval (a, b) contains an interval (c, d)

disjoint from A_k . All other sets are of second category.⁶ Although interest in the specific problems that inspired Baire has waned, the importance of his work has not. The subject of functional analysis rests on three main theorems. Two of them are direct consequences of what is called the Baire category theorem, which asserts that a complete metric space is of second category as a subset of itself. These fundamental pillars of functional analysis—the closed graph theorem and the open mapping theorem—cannot be proved without the Baire category theorem. Here we have an example of an unintended and fortuitous consequence, in which a result turned out to be useful in an area not considered by its discoverer.

PROBLEMS AND QUESTIONS

Mathematical Problems

43.1. On the basis of the geometric series

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \dots$$

Euler was willing to say that $1 - 5 + 25 - 125 + \dots = \frac{1}{1+5} = \frac{1}{6}$. Later analysts rejected this interpretation of infinite series and confined themselves to series that converge in the ordinary sense. (Such a series cannot converge unless its general term tends to zero.) Kurt Hensel (1861–1941), showed in 1905 that it is possible to define a notion of distance, the *p-adic metric*, by saying that an integer is close to zero if it is divisible by a large power of the prime number p (in the present case $p = 5$). Specifically, the distance from m to 0 is given by $d(m, 0) = 5^{-k}$, where 5^k divides m but 5^{k+1} does not divide m . The distance between 0 and the rational number $r = m/n$ is then by definition $d(m, 0)/d(n, 0)$. Show that $d(1, 0) = 1$. Show that, if the distance between two rational numbers r and s is defined to be $d(r - s, 0)$, then in fact the series just mentioned does converge to $\frac{1}{6}$ in the sense that $d(S_n, \frac{1}{6}) \rightarrow 0$, where S_n is the n th partial sum.

43.2. Consider the functions

$$f_n(x) = \begin{cases} n^2x, & \text{if } 0 \leq x \leq \frac{1}{n}, \\ 2n - n^2x, & \text{if } \frac{1}{n} \leq x \leq \frac{2}{n}, \\ 0, & \text{if } \frac{2}{n} \leq x \leq 1. \end{cases}$$

Show that $f_n(x) \rightarrow 0$ as $n \rightarrow \infty$ for each x satisfying $0 \leq x \leq 1$, but $\int_0^1 f_n(x) dx = 1$ for all n . Why does this sequence not satisfy the hypotheses of the Lebesgue dominated convergence theorem?

⁶The sets A_k are said to be *nowhere dense*. In his famous treatise *Mengenlehre (Set Theory)*, Felix Hausdorff (1862–1942) criticized the phrase *of first category* as “colorless.”

- 43.3.** Show that on a finite interval $[a, b]$ the space $L_p([a, b])$ is contained in $L_q([a, b])$ if $q < p$, while the opposite is true for the l_p spaces of sequences. (The space l_p consists of all sequences $\{a_n\}_{n=1}^{\infty}$ such that $\sum_{n=1}^{\infty} |a_n|^p < \infty$.) Which, if either, of these statements is true for the interval $[0, \infty)$?

Historical Questions

- 43.4.** Why was the concept of uniform convergence important in the application of the principles of analysis to infinite series and integrals? Why was it too restrictive for the needs of modern analysis?
- 43.5.** How does the Lebesgue integral differ from the Riemann integral, and why is the latter inadequate for the needs of modern analysis?
- 43.6.** What is the Baire category theorem, and why is it important in modern analysis?

Questions for Reflection

- 43.7.** What are the advantages, if any, of building a theory by starting with abstract definitions, then later proving a structure theorem showing that the abstract objects so defined are really composed of familiar simple objects? (Recall that Cauchy preferred to begin his discussion of analytic functions with the abstract property of differentiability, while Weierstrass preferred the more concrete definition of an analytic function as a power series. But it turns out that the two classes are the same.)
- 43.8.** Why did the naive application of finite rules to infinite series lead to paradoxes?
- 43.9.** One consequence of the Lebesgue dominated convergence theorem is that if a uniformly bounded sequence of continuous functions $f_n(x)$ tends to 0 at each point $x \in [0, 1]$, then $\int_0^1 f_n(x) dx \rightarrow 0$. This theorem is quite difficult to prove in the context of the Riemann integral, but becomes a trivial consequence of a basic result when the integrals are interpreted as Lebesgue integrals. What advantage does this fact point to when the Riemann integral is compared with the Lebesgue integral?

Set Theory

Set theory is the common language now used in all areas of mathematics. Because it is the language everyone writes in, it is difficult to imagine a time when mathematicians did not use the word *set* or think of sets of points. Yet that time is not long past; it was less than 150 years ago. Before that time, mathematicians spoke of geometric figures, or they spoke of points and numbers having certain properties, without thinking of those points and numbers as being assembled in a set.

44.1. TECHNICAL BACKGROUND

Although the founder of set theory, Georg Cantor (1845–1918), was motivated by both geometry and analysis, for reasons of space we shall discuss only the analytic connection, which was the more immediate one. It is necessary to introduce some technical details in order to explain how a problem in analysis leads to the general notion of a set and an ordinal number. We begin with the topic that Riemann developed for his 1854 lecture but did not use because Gauss preferred his geometric lecture. That topic was uniqueness of trigonometric series, and it was published in 1867, the year after Riemann's death. Riemann aimed at proving that if a trigonometric series converged to zero at every point, all of its coefficients were zero. That is,

$$\frac{1}{2}a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx) \equiv 0 \implies a_n = 0 = b_n.$$

Riemann assumed that the coefficients a_n and b_n tend to zero, saying that it was clear to him that without that assumption, the series could converge only at isolated points.¹ In order to prove his uniqueness theorem, Riemann integrated twice to form the continuous function

$$F(x) = Ax + B + \frac{1}{4}a_0x^2 - \sum_{n=1}^{\infty} \frac{(a_n \cos nx + b_n \sin nx)}{n^2}.$$

¹Leopold Kronecker (1823–1891) pointed out later that this assumption could be omitted. Cantor showed, as Riemann implied, that it was deducible from the mere convergence of the series.

His object was to show that $F(x)$ must be a linear function, so that $G(x) = F(x) - Ax - B - \frac{1}{4}a_0x^2$ would be a quadratic polynomial that was also periodic, and hence itself a constant, from which it would follow, first that $a_0 = 0 = A$, and then that all the other a_n and all the b_n are zero. To that end, he showed that its generalized second derivative

$$F_g''(x) = \lim_{h \rightarrow 0} \frac{F(x+h) + F(x-h) - 2F(x)}{h^2}$$

was zero wherever the original series converged to zero.² Riemann proved that in any case

$$\lim_{h \rightarrow 0} \frac{(F(x+h) - F(x)) + (F(x-h) - F(x))}{h} = 0.$$

The important implication of this last result is that *the function $F(x)$ cannot have a corner*. If it has a right-hand derivative at a point, it also has a left-hand derivative at the point, and the two one-sided derivatives are equal. This fact was a key step in Cantor's work.

44.2. CANTOR'S WORK ON TRIGONOMETRIC SERIES

In 1872, Cantor published his first paper on uniqueness of trigonometric series, finishing the proof that Riemann had set out to give: that a trigonometric series that converges to zero at every point must have all its coefficients equal to zero. In following the program of proving that $F(x)$ is linear and hence constant, he observed that it was not necessary to assume that the series converged to zero at every point. A finite number of exceptions could be allowed, at which the series either diverged or converged to a nonzero value. For $F(x)$ is certainly continuous, and if it is linear on $[a, b]$ and also on $[b, c]$, the fact that it has no corners implies that it must be linear on $[a, c]$. Hence any isolated exceptional point b could be discounted.

The question therefore naturally arose: Can one allow an infinite number of exceptional points? Here one comes up against a theorem known as the Bolzano–Weierstrass theorem, which asserts that the exceptional points cannot all be isolated. They must have at least one point of accumulation.³ But exceptional points isolated from other exceptional points could be discounted, just as before. That left only their points of accumulation. If these were isolated—in particular, if there were only finitely many of them—the no-corners principle would once again imply uniqueness of the series.

44.2.1. Ordinal Numbers

Cantor saw the obvious induction. Denoting the set of points of accumulation of a set P (what we now call the derived set) by P' , he knew that $P' \supseteq P'' \supseteq P''' \supseteq \dots$. Thus, if at some finite term in this non-increasing sequence of sets a finite set was obtained, the uniqueness

²Hermann Amandus Schwarz (1843–1921) later showed that if $F(x)$ is continuous on a closed interval $[a, b]$ and $F_g''(x) \equiv 0$ on the open interval (a, b) , then $F(x)$ is linear on the closed interval $[a, b]$.

³A point of accumulation of a set is a point, every neighborhood of which contains infinitely many points of the set. It is also called a *cluster point* and (confusingly and more commonly) a *limit point*.

theorem would remain valid. But the study of these sets of points of accumulation turned out to be even more interesting than trigonometric series themselves. No longer dealing with geometrically regular sets, Cantor was delving into point-set topology, as we now call it. No properties of a geometric nature were posited for the exceptional points he was considering, beyond the assumption that the sequence of derived sets contains a finite set as one of its terms (and all subsequent sets in the sequence will be empty). Although the points of any particular set might be easily describable, Cantor needed to discuss the general case. He needed the abstract concept of “set-hood.” Cantor felt compelled to dig to the bottom of this matter and soon abandoned trigonometric series to write a series of papers on “infinite linear point-manifolds.”⁴

Cantor noticed the possibility of defining the derived sets of transfinite order. If the n th-level derived set is $P^{(n)}$, the nesting of these sets allows the natural definition of the derived set of infinite order $P^{(\infty)}$ as the intersection of all sets of finite order. But then one could consider derived sets even at the transfinite level: the derived set of $P^{(\infty)}$ could be defined as $P^{(\infty+1)} = (P^{(\infty)})'$. Cantor had discovered the infinite ordinal numbers. He did not at first recognize them as numbers, but rather regarded them as “symbols of infinity” (see Ferreirós, 1995).

44.2.2. Cardinal Numbers

Cantor was not only an analyst. He had written his dissertation under Kronecker and Ernst Eduard Kummer (1810–1893) on number-theoretic questions. Only two years after he wrote his first paper in trigonometric series, he noticed that his set-theoretic principles led to another interesting conclusion. *Transcendental numbers exist*. The set of algebraic numbers is a countable set (as we would now say, in the familiar language that we owe to Cantor), but the set of real numbers is not. Cantor had proved this point to his satisfaction in a series of exchanges of letters with Dedekind.

There are two versions of this proof, one due to Cantor and one due to Dedekind, but both involve getting nested sequences of closed intervals that exclude, one at a time, the elements of any given sequence $\{a_n\}$ of numbers. The intersection of the intervals must then contain a number not in the sequence. In his private speculations some 40 years later, Luzin noted that Cantor was actually assuming more than the mere *existence* of the countable set $\{a_n\}$. In order to construct a point not in it, one had to know something about each of its elements, enough to find a subinterval of the previous closed interval that would exclude the next element. On that basis, he concluded that Cantor had proved that there was no *effective enumeration* of the reals, not that the reals were uncountable. Luzin thus raised the question of what it could mean for an enumeration to “exist” if it was not effective. He too delved into philosophy to find out the meaning of “existence.”

By showing in a seemingly constructive way how, given any countable enumeration of real numbers, one can exhibit a real number not in the list, Cantor had shown (he thought) that there must exist transcendental numbers. Given the complicated constructions of such numbers by Liouville and the difficulty of the proofs by Hermite and Lindemann that e and

⁴The word *manifold* (*Mannigfaltigkeit*) does not denote the geometric object now called a manifold. It means a *multitude*, which is also the earlier meaning of the English term. The modern German word for a set is *Menge*, also used in the phrase *eine Menge von...* to mean *lots of...* Russian mathematicians followed this German usage and used the word *mnozhestvo* (multitude) to denote a set. The French use the word *ensemble*, which has the same connotations as the English word.

π are transcendental, this concise proof of their abstract existence seemed to have merit. The property Cantor relied on in this proof led to the concept of a cardinal number, two sets being of the same cardinality if they could be placed in one-to-one correspondence. To establish such correspondences, Cantor allowed certain methods of defining sets and functions that went beyond what mathematicians had been used to seeing. The result was a controversy that lasted some two decades.

Grattan-Guinness (2000, p. 125) has pointed out that Cantor emphasized five different aspects of point sets: their topology, dimension, measure, cardinality, and ordering. In the end, point-set topology was to become its own subject, and dimension theory became part of both algebraic and point-set topology. Measure theory became an important part of modern integration theory and had equally important applications to the theory of probability and random variables. Cardinality and ordering remained as an essential core of set theory, and the study of sets in relation to their complexity rather than their size became known as *descriptive set theory*.

Although descriptive set theory produces its own questions, it had at first a close relation to measure theory, since descriptive set theory was needed to specify which sets could be measured. Borel was conservative, allowing that the kinds of sets one could clearly define would have to be obtained by a *finite* sequence of operations, each of which was either a countable union or a countable intersection or a complementation, starting from ordinary open and closed sets. Ultimately those of a less constructive disposition than Borel honored him with the creation of the *Borel sets*, which is the smallest class that contains all closed subsets and also contains the complement of any of its sets and the union of any countable collection of its sets. This class, now called a σ -algebra by analysts and a σ -field by probabilists, can be “constructed” only by a transfinite induction.

Set theory, while aiming to provide a foundation of clear and simple principles for all of mathematics, soon threw up its own unanswered mathematical questions. The most prominent of these was the continuum question. Cantor had shown that the set of all real numbers could be placed in one-to-one correspondence with the set of all subsets of the integers. He denoted the cardinality of the integers as \aleph_0 and the cardinality of the real numbers as \mathfrak{c} (where \mathfrak{c} stands for “continuum”). The question naturally arose whether there was any subset of the real numbers that had a cardinality between these two. Cantor struggled for a long time to settle this issue. One major theorem of set theory, known as the *Cantor–Bendixson theorem*,⁵ after Ivar Bendixson (1861–1935), asserts that every closed set is the union of a countable set and a perfect set (a set that is equal to its derived set). Since it is easily proved that a nonempty perfect subset of the real numbers has cardinality \mathfrak{c} , it follows that every uncountable closed set contains a subset of cardinality \mathfrak{c} . Thus a set of real numbers having cardinality between \aleph_0 and \mathfrak{c} cannot be a closed set.

Many mathematicians, especially the Moscow mathematicians after the arrival of Luzin as professor in 1915, worked on this problem. Luzin’s students Pavel Sergeevich Aleksandrov (1896–1982) and Mikhail Yakovlevich Suslin (1894–1919) proved that any uncountable Borel set must contain a nonempty perfect subset and thus must have cardinality \mathfrak{c} . Indeed, they proved this fact for a slightly larger class of sets called *analytic sets*. Luzin

⁵Ferreirós (1995) points out that it was the desire to prove this theorem adequately, in 1882, that really led Cantor to treat transfinite ordinal numbers as numbers. He was helped toward this discovery by Dedekind’s pointing out to him the need to use finite ordinal numbers to define finite cardinal numbers.

then proved that a set was a Borel set if and only if the set and its complement were both analytic sets.

The problem of the continuum remained open until 1938, when Kurt Gödel (1906–1978) partially closed it by showing that set theory is consistent with the continuum hypothesis and the axiom of choice,⁶ provided that it is consistent without them. Closure came to this question in 1963, when Paul Cohen (1934–2007)—like Cantor, he began his career by studying uniqueness of trigonometric series representations—showed that the continuum hypothesis and the axiom of choice are independent of the other axioms of set theory.

44.3. THE RECEPTION OF SET THEORY

Some mathematicians believed that set theory was an unwarranted intrusion of philosophy into mathematics. One of those was Cantor's teacher Leopold Kronecker. Although Cantor was willing to regard the existence of transcendental numbers as having been proved just because the real numbers were “too numerous” to be exhausted by the algebraic numbers, Kronecker preferred a more constructivist approach. His most famous utterance,⁷ and one of the most famous in the history of mathematics, is: “The good Lord made the integers; everything else is a human creation.” (“*Die ganzen Zahlen hat der liebe Gott gemacht; alles andere ist Menschenwerk.*”) That is, the only infinity he admitted was the series of positive integers $1, 2, \dots$. Beyond that point, everything was human-made and therefore had to be finite. If you spoke of a number or function, you had an obligation to say how it was defined. His 1845 dissertation, which he was unable to polish to his satisfaction until 1881, when he published it as “Foundations of an arithmetical theory of algebraic quantities” in honor of his teacher Kummer, shows how conservative he was in his definitions. Instead of an arbitrary *field* defined axiomatically as we would now do, he wrote:

A domain of rationality is in general an *arbitrarily* bounded domain of magnitudes, but only to the extent that the concept permits. To be specific, since a domain of rationality can be enlarged only by the adjoining of arbitrarily chosen elements \mathfrak{R} , each arbitrary extension of its boundary requires the simultaneous inclusion of *all* quantities rationally expressible in terms of the new element.

In this way, while one could enlarge a field to make an equation solvable, the individual elements of the larger field could still be described constructively. Kronecker's concept of a general field can be described as “finitistic.” It is the minimal object that contains the elements necessary to allow arithmetic operations. Borel took this point of view in regard to measurable sets, and Hilbert was later to take a similar point of view in describing formal languages, saying that a meaningful formula must be obtained from a specified list of elements by a finite number of applications of certain rules of combination. This approach was safer and more explicit than, for example, John Bernoulli's original definition of a function as an expression formed “in some manner” from variables and constants. The “manner” was limited in a very definite way.

⁶Gödel actually included four additional assumptions in his consistency proof, one of the other two being that there exists a set that is analytic but is not a Borel set.

⁷He made this statement at a meeting in Berlin in 1886 (see Grattan-Guinness, 2000, p. 122).

44.3.1. Cantor and Kronecker

Cantor believed that Kronecker had conspired against him to delay the publication of his first paper on infinite cardinal numbers. Whether that is the case or not, it is clear that Kronecker would not have approved of some of Cantor's principles of inference. As Grattan-Guinness points out, much of what is believed about the animosity between Cantor and Kronecker is based on Cantor's own reports, which may be unreliable. Cantor was subject to periodic bouts of depression, probably caused by metabolic imbalances having nothing to do with his external circumstances. In fact, he had little to complain of in terms of the acceptance of his theories. It is true that there was some resistance to it, notably from Kronecker (until his death in 1891) and then from Poincaré. But there was also a great deal of support, from Weierstrass, Klein, Hilbert, and many others. In fact, as early as 1892, the journal *Bibliotheca mathematica* published a "Notice historique" on set theory by Giulio Vivanti (1859–1949), mentioning that there had already been several expositions of the theory, and that it was still being developed by mathematicians, applied to the theory of functions of a real variable, and studied from a philosophical point of view.

44.4. EXISTENCE AND THE AXIOM OF CHOICE

In the early days, Cantor's set theory seemed to allow a remarkable amount of freedom in the "construction" or, rather, the conjuring into existence, of new sets. Cantor seems to have been influenced in his introduction of the term *set* by an essay that Dedekind began in 1872, but did not publish until 1887 (see Grattan-Guinness, 2000, p. 104), in which he referred to a "system" as "various things a, b, c, \dots comprehended from any cause under one point of view." Dedekind defined a "thing" to be "any object of our thought." Just as Descartes was able to conceive many things clearly and distinctly, mathematicians seemed to be able to form many "things" into "systems." For example, given *any* set A , one could conceive of another set whose members were the subsets of A . This set is nowadays denoted 2^A and called the *power set* of A . If A has a finite number n of elements, then 2^A has 2^n elements, counting the improper subsets \emptyset and A .

It was not long, however, before the indiscriminate use of this freedom to form sets led to paradoxes. The most famous of these is Russell's paradox, which will be discussed in the next chapter. In modern set theory, this paradox is avoided by distinguishing between a set, which is a class that may or may not have members but at least is itself a member of some other class, and a proper class, which has members, but is not itself the member of any class. Admission to the elite company of sets is carefully controlled by the axioms. The empty class is declared to be a set by fiat. Other sets arise from operations on classes known to be sets.

One source of the paradoxes is that "existence" has a specialized mathematical meaning in set theory, which has the consequence that much of the action in a proof takes place "offstage." That is, certain objects needed in a proof are simply declared to exist by saying, "Let there be. . .," but no procedure for constructing them is given. Proofs relying on the abstract existence of such objects, when it is not possible to choose a particular object and examine it, became more and more common in the twentieth century. Indeed, much of measure theory, topology, and functional analysis would be impossible without such proofs. The principle behind these proofs later came to be known as *Zermelo's axiom*, after Ernst Zermelo (1871–1953), who first formulated it in 1904 to prove that every set could be well

ordered.⁸ It was also known as the *principle of free choice* (in German, *Auswahlprinzip*) or, more commonly in English, the *axiom of choice*. In its broadest form this axiom states that *there exists a function f defined on the class of all nonempty sets such that $f(A) \in A$ for every nonempty set A* . Intuitively, if A is nonempty, there exist elements of A , and $f(A)$ chooses an element from every nonempty set A .

This axiom is used in many proofs. Probably the earliest (see Moore, 1982, p. 9) is Cantor's proof that a countable union of countable sets is countable. The proof goes as follows. Assume that A_1, A_2, \dots are countable sets, and let $A = A_1 \cup A_2 \cup \dots$. Then A is countable. For, let the sets A_j be enumerated, as follows:

$$\begin{aligned} A_1 &= a_{11}, a_{12}, \dots, \\ A_2 &= a_{21}, a_{22}, \dots, \\ &\vdots \\ A_n &= a_{n1}, a_{n2}, \dots, \\ &\vdots \end{aligned}$$

Then the elements of A can be enumerated as follows: $a_{11}, a_{12}, a_{21}, a_{13}, a_{22}, a_{31}, \dots$, where the elements whose ranks are larger than the triangular number $T_n = n(n+1)/2$ but not larger than $T_{n+1} = (n+1)(n+2)/2$ are those for which the sum of the subscripts is $n+2$. There are $n+1$ such elements and $n+1$ such ranks. It is a very subtle point to notice that this proof assumes more than the mere existence of an enumeration of *each* of the sets, which is given in the hypothesis. It assumes the *simultaneous* existence of infinitely many enumerations, one for each set. The reasoning appears to be so natural that one would hardly question it. If a real choice exists at each stage of the proof, why can we not assume that infinitely many such choices have been made? As Moore notes, without the axiom of choice, it is consistent to assume that the real numbers can be expressed as a countable union of countable sets.⁹

Zermelo made this axiom explicit and showed its connection with ordinal numbers. The problem then was either to justify the axiom of choice, or to find a more intuitively acceptable substitute for it, or to find ways of doing without such "noneffective" concepts. A debate about this axiom took place in 1905 in the pages of the *Comptes rendus* of the Paris Academy of Sciences, which published a number of letters exchanged among Borel, Lebesgue, Baire, and Jacques Hadamard (1865–1963).¹⁰ Borel had raised objections to Zermelo's proof that every set could be well-ordered on the grounds that it assumed an infinite number of enumerations. Hadamard thought it an important distinction that in some cases the enumerations were all independent, as in Cantor's proof above, but in others each depended for its definition on other enumerations having been made in correspondence with a smaller ordinal number. He agreed that the latter should not be used transfinitely. Borel

⁸A set is *well-ordered* if any two elements can be compared and every nonempty subset has a smallest element. The positive integers are well ordered by the usual ordering. The positive real numbers are not, since there is no smallest positive number.

⁹Not *every* countable union of countable sets is uncountable, however; the rational numbers remain countable, because an explicit counting function can be constructed.

¹⁰These letters were translated into English and published by Moore (1982, pp. 311–320).

had objected to using the axiom of choice nondenumeratively, but Hadamard thought that this usage brought no further damage, once a denumerable infinity of choices was allowed. He also mentioned the distinction due to Jules Tannery (1848–1910) between *describing* an object and *defining* it. To Hadamard, describing an object was a stronger requirement than defining it. To supply an example for him, we might mention a well-ordering of the real numbers, which is *defined* by the phrase itself, but effectively *indescribable*. Hadamard noted Borel's own work on analytic continuation and pointed out how it would change if the only power series admitted were those that could be effectively described. The difference, he said, belongs to psychology, not mathematics.

Hadamard received a response from Baire, who took an even more conservative position than Borel. He said that once an infinite set was spoken of, "the comparison, *conscious or unconscious*, with a bag of marbles passed from hand to hand must disappear completely."¹¹ The heart of Baire's objection was Zermelo's *supposition* that to each (nonempty) subset of a set M there corresponds one of its elements." As Baire said, "all that it proves, as far as I am concerned, is that we do not perceive a contradiction" in imagining any set well-ordered.

Responding to Borel's request for his opinion, Lebesgue gave it. He said that Zermelo had very ingeniously shown how to solve problem A (to well-order any set) provided one could solve problem B (to choose an element from every nonempty subset of a given set). He remarked, probably with some irony, that, "Unfortunately, problem B is not easy to resolve, it seems, except for the sets that we know how to well-order." Lebesgue mentioned a concept that was to play a large role in debates over set theory, that of "effectiveness," roughly what we would call constructibility. He interpreted Zermelo's claim as the assertion that a well-ordering *exists* (that word again!) and asked a question, which he said was "hardly new": *Can one prove the existence of a mathematical object without defining it?* One would think not, although Zermelo had apparently proved the existence of a well-ordering (and Cantor had proved the existence of a transcendental number) without *describing* it. Lebesgue and Borel preferred the verb *to name* (*nommer*) when referring to an object that was defined effectively, through a finite number of uses of well-defined operations on a given set of primitive objects.

After reading Lebesgue's opinion, Hadamard was sure that the essential distinction was between what is determined and what is described. He compared the situation with the earlier debate over the allowable definitions of a function. But, he said, uniqueness was not an issue. If one could say "For each x , there exists a number satisfying. . . . Let y be this number," surely one could also say "For each x , there exists an infinity of numbers satisfying. . . . Let y be one of these numbers." In that statement, he had put his finger squarely on one of the paradoxes of set theory (the Burali-Forti paradox, discussed in the next section). "It is the very existence of the set W that leads to a contradiction. . . the general definition of the word *set* is incorrectly applied." What *is* the definition of the word *set*?

The validity and value of the axiom of choice remained a puzzle for some time. It leads to short proofs of many theorems whose statements are constructive. For example, it proves the existence of a nonzero translation-invariant Borel measure on any locally compact abelian group. Since such a measure is provably unique (up to a constant multiple), there ought to be effective proofs of its existence that do not use the axiom of choice (and indeed there are). One benefit of the 1905 debate was a clarification of equivalent forms of the axiom of choice

¹¹Luzin said essentially the same in his journal: "What makes the axiom of choice seem reasonable is the picture of reaching into a set and helping yourself to an element of it."

and an increased awareness of the many places where it was being used. A list of important theorems whose proof used the axiom was compiled for Luzin's seminar in Moscow in 1918. The list showed, as Luzin wrote in his journal, that "almost nothing is proved without it." Luzin was horrified, and spent some restless nights pondering the situation.

The axiom of choice is ubiquitous in modern analysis; little would remain of functional analysis or point-set topology if it were omitted entirely, although weaker assumptions might suffice. The Baire category theorem mentioned in the previous chapter cannot be proved without it. It is fortunate, therefore, that its consistency with, and independence of, the other axioms of set theory has been proved. The consequences of this axiom are suspiciously strong. In 1924 Alfred Tarski (1901–1983) and Stefan Banach (1892–1945) deduced from it that any two sets A and B in ordinary three-dimensional Euclidean space, each of which contains some ball, can be decomposed into pairwise congruent subsets. This means, for example, that a cube the size of a grain of salt (set A) and a ball the size of the sun (set B) can be written as disjoint unions of sets A_1, \dots, A_n and B_1, \dots, B_n respectively such that A_i is congruent to B_i for each i . This result (the Banach–Tarski paradox) is very difficult to accept. It can be rationalized only by realizing that the notion of existence in mathematics has no metaphysical content. To say that the subsets A_i, B_i "exist" means only that a certain formal statement beginning $\exists \dots$ is deducible from the axioms of set theory.

PROBLEMS AND QUESTIONS

Mathematical Problems

- 44.1.** Bertrand Russell pointed out that some applications of the axiom of choice are easier to avoid than others. For instance, given an infinite collection of pairs of shoes, describe a way of choosing one shoe from each pair. Could you do the same for an infinite set of pairs of socks?
- 44.2.** Let $k = k(m)$ denote the largest integer less than the number

$$\frac{\sqrt{8m+1}-1}{2}.$$

(The greatest integer less than a positive real number x can be described as $[x] - [1 - x + [x]]$, where $[x]$ is the standard greatest-integer function.) Show that the mapping

$$f(m) = \frac{k^2 + 3k + 4 - 2m}{2m - k(k+1)}$$

is an enumeration of all positive rational numbers and in fact, each positive rational number occurs an infinite number of times in this enumeration. (Show that the number $\frac{p}{q}$ occurs as $f(n)$, where $n = 1 + \frac{1}{2}((p+q)^2 - 3p - q)$. Thus, for example, $\frac{3}{2}$ occurs at $n = 8$ and its reciprocal $\frac{2}{3}$ at $n = 9$.)

- 44.3.** According to the axioms of set theory, every set can be well-ordered. By the continuum hypothesis, there is a one-to-one correspondence $x \leftrightarrow \xi$ between the real numbers x such that $0 \leq x \leq 1$ and the countable ordinal numbers ξ . Suppose such a correspondence is given and we define a function $f(x, y)$ on pairs of real numbers

between 0 and 1 by specifying that $f(x, y) = 1$ if $\xi < \nu$, where $x \leftrightarrow \xi$ and $y \leftrightarrow \nu$, and $f(x, y) = 0$ otherwise. Show that

$$\int_0^1 \int_0^1 f(x, y) dy dx = 1, \quad \int_0^1 \int_0^1 f(x, y) dx dy = 0.$$

Why does this result not contradict the principal from calculus whereby a double integral can be evaluated as an iterated integral in either order. In Lebesgue integration, this principle that a double integral can be evaluated as an iterated integral still holds and is known as Fubini's theorem, after Guido Fubini (1879–1943).

Historical Questions

- 44.4. What was Cantor's original motive for studying sets of points on the real line?
- 44.5. On what philosophical grounds did Kronecker object to Cantor's methods?
- 44.6. What consensus did the mathematical community reach regarding set theory, especially the axiom of choice?

Questions for Reflection

- 44.7. In what sense do mathematical entities exist? When we prove the existence of a root of an equation or a minimal curve having a certain property without exhibiting it explicitly in terms of more familiar mathematical entities, what is the effective, practical meaning of that proof?
- 44.8. What does the Banach–Tarski paradox suggest about the meaning of mathematical concepts and their application to the physical world?
- 44.9. The philosopher Immanuel Kant (1724–1804) described mathematical knowledge (of arithmetic and geometry) as something innate, what he called *synthetic, a priori* knowledge. It was synthetic because propositions like the equality $5 + 7 = 12$ did not follow from pure logic, that is, the definition of the concepts of 5, 7, 12, and addition. It was *a priori* because it was not learned from experience. We “just know” it; it's part of the way our brains are wired, as we would now say. This view, which possibly Kronecker would have agreed with, conflicts with the aims of set theory, in which arithmetic is to be derived logically from more primitive concepts. Where would you place the rock-bottom foundation layer of mathematics: in the simple arithmetic of the positive integers (with Kant and Kronecker), or in the intuitive idea of a set or class (with Cantor, Hilbert, and others)? Or is there a third possibility that seems better to you?

Logic

The mathematization of logic has a prehistory that goes back to Leibniz (not published in his lifetime), but we shall focus on mostly the nineteenth-century work. After a brief discussion of the preceding period, we examine the period from 1847 to 1931. This period opens with the treatises of Boole and De Morgan and closes with Gödel's famous incompleteness theorem. Our discussion is not purely about logic in the earlier parts, since the earlier writers considered both logical and probabilistic reasoning.

45.1. FROM ALGEBRA TO LOGIC

Leibniz was one of the first to conceive the idea of creating an artificial language in which to express propositions. He compared formal logic to the lines drawn in geometry as guides to thought. If the language encoded thought accurately, thought could be analyzed in a purely mechanical manner:

“... when disagreements arise, there will be no more need for two philosophers to argue than for two accountants to do so. For it will suffice for them to take pen in hand, sit at their counting-boards, and say to each other, ‘Let us calculate.’ [Gerhardt, 1971, Bd. 7, p. 200]

In another place he wrote:

Ordinary languages, though mostly helpful for the inferences of thought, are yet subject to countless ambiguities and cannot do the task of a calculus, which is to expose mistakes in inference... This remarkable advantage is afforded up to date only by the symbols of arithmeticians and algebraists, for whom inference consists only in the use of characters, and a mistake in thought and in the calculus is identical. [Quoted by Bochenski, 1961, p. 275]

Thus did Leibniz begin to extend the application of Euclidean formalism to all of philosophy. But it was not the Euclidean axioms that led him to do this. It was the spectacular development of algebra in his own time. This transition from the visual to the verbal, from intuition to language, was one of the prominent features of modern mathematics. It works quite well in arithmetic and other discrete systems and fairly well even with continuous systems such as the real numbers and the objects of geometry.

The ideal enunciated by Leibniz remains largely unfulfilled when it comes to settling philosophical disagreements. It reflects an oversimplified and optimistic view of human beings as basically rational creatures. This sort of optimism continued into the early nineteenth century, as exemplified by the *Handbook of Political Fallacies* by the philosopher Jeremy Bentham (1748–1832). But while the complex questions of the world of nature and society could not be mastered through logic alone, mathematics proved more amenable to the influence of logic. The influence, however, was bidirectional. In fact, there is a paradox, if one thinks of logic as being the rudder that steers mathematical arguments and keeps them from going astray. As the American philosopher/mathematician Charles Sanders Peirce (1839–1914) wrote in 1896, reviewing a book on logic:

It is a remarkable historical fact that there is a branch of science in which there has never been a prolonged dispute concerning the proper objects of that science. It is mathematics. . . Hence, we homely thinkers believe that, considering the immense amount of disputation there has always been concerning the doctrines of logic, and especially concerning those which would otherwise be applicable to settle disputes concerning the accuracy of reasonings in metaphysics, the safest way is to appeal for our logical principles to the science of mathematics. [Quoted in Bochenski, 1961, pp. 279–280]

Peirce seemed to believe that, far from sorting out the mathematicians, logicians should turn to them for guidance. But we may dispute his assertion that there has never been a prolonged dispute about the proper objects of mathematics. Zeno's paradoxes concern that very question. In Peirce's own day, Kronecker and Cantor were at opposite ends of a dispute about what is and is not proper mathematics, and that discussion continues, politely, down to the present day. See, for example, the book (1997) by Reuben Hersh (b. 1927).

Leibniz noted in the passage quoted above that algebra had the advantage of a precise symbolic language, which he held up as an ideal for clarity of communication. Algebra was one of the sources of mathematical logic. When De Morgan translated a French algebra textbook into English in 1828, he defined algebra as "the part of mathematics in which symbols are employed to abridge and generalize the reasonings which occur in questions relating to numbers." Thus, for De Morgan at the time, the symbols represented numbers, but *unspecified* numbers, so that reasoning about them applied to any particular numbers. Algebra was a ship anchored in numbers, but it was about to slip its anchor.

Only two years later (in 1830) George Peacock (1791–1858) wrote a treatise on algebra in which he proposed that algebra be a purely symbolic science independent of any arithmetical interpretation. This step was a radical innovation at the time, considering that abstract groups, for example, were not to appear for several more decades. The assertion that the formula $(a - b)(a + b) = a^2 - b^2$ holds independently of any numerical values that replace a and b , for example, almost amounts to an axiomatic approach to mathematics. De Morgan's ideas on this subject matured during the 1830s, and at the end of the decade he wrote:

When we wish to give the idea of symbolical algebra. . . we ask, firstly, what symbols shall be used (without any reference to meaning); next, what shall be the laws under which such symbols are to be operated upon; the deduction of all subsequent consequences is again an application of common logic. Lastly, we explain the meanings which must be attached to the symbols, in order that they may have prototypes of which the assigned laws of operation are true. [Quoted by Richards, 1987, pp. 15–16]

This set of procedures is still the way in which mathematical logic operates, although the laws under which the symbols are to be operated on are now more abstract than De Morgan probably had in mind. To build a formal language, you first specify which sequences of symbols are to be considered “well-formed formulas,” that is, formulas capable of being true or false. The criterion for being well-formed must be purely formal, one that could in principle be used by a computer. Next, the sequences of well-formed formulas that are to be considered deductions are specified, again purely formally. The *syntax* of the language is specified by these two sets of rules, and the final piece of the construction, as De Morgan notes, is to specify its *semantics*, that is, the interpretation of its symbols and formulas. Here again, the modern world takes a more formal and abstract view of “interpretation” than De Morgan probably intended. For example, the semantics of propositional calculus consists of truth tables. After specifying the semantics, one can ask such questions as whether the language is consistent (incapable of proving a false proposition), complete (capable of proving all true propositions), or categorical (allowing only one interpretation, up to isomorphism).

In his 1847 treatise *Formal Logic*, De Morgan went further, arguing that “we have power to invent new meanings for all the forms of inference, in every way in which we have power to make new meanings of *is* and *is not*. . . .” This focus on the meaning of *is* was very much to the point. One of the disputes that Peirce overlooked in the quotation above is the question of what principles allow us to infer that an object “exists” in mathematics. We have seen this question in the eighteenth-century disagreement over what principles are allowed to define a function. In the case of symbolic algebra, where the symbols originally represented numbers, the existence question was still not settled to everyone’s liking in the early nineteenth century. That is why Gauss stated the fundamental theorem of algebra in terms of real factorizations alone. Here De Morgan was declaring the right to create mathematical entities by *fiat*, subject to certain restrictions. That enigmatic “exists” is indispensable in first-order logic, where the negation of “For every x , P ” is “For some x , not- P .” But what can “some” mean unless there actually *exist* objects x ? This defect was to be remedied by De Morgan’s friend George Boole (1815–1864).

In De Morgan’s formal logic, this “exists” remains hidden: When he talks about a class X , it necessarily has members. Without this assumption, even his very first example is not a valid inference. He gives the following table by way of introduction to the symbolic logic that he is about to introduce:

<i>Instead of:</i>	<i>Write:</i>
All men will die	Every Y is X
All men are rational beings	Every Y is Z
Therefore some rational beings will die	Therefore some Z s are X ’s

De Morgan’s notation in this work was not the best, and very little of it has caught on. He used a parenthesis in roughly the same way as the modern notation for implication. For example, $X \supset Y$ denoted the proposition “Every X is a Y .” Nowadays we would write $X \supset Y$ (read “ X horseshoe Y ”) for “ X implies Y .” (To add to the confusion, if x is the set of objects for which X is true and y the set for which Y is true, then $X \supset Y$ actually means $x \subset y$.) The rest of his notation— $X : Y$ for “Some X ’s are not Y s,” $X.Y$ for “No X ’s are Y s,” and $X Y$ for “Some X ’s are Y s”—is no longer used. For the negation of these properties he used lowercase letters, so that x denoted not- X . De Morgan introduced the useful “necessary”

and “sufficient” language into implications: $X \supset Y$ meant that Y was *necessary* for X and X was *sufficient* for Y . He gave a table of the relations between X or x and Y or y for the relations $X \supset Y$, $X.Y$, $Y \supset X$, and $x . y$. For example, given that X implies Y , he noted that this relation made Y necessary for X , y an impossible condition for X , y a sufficient condition for x , and Y a contingent (not necessary, not sufficient, not impossible) condition for x .

For compound propositions, he wrote PQ for conjunction (his word), meaning both P and Q are asserted, and P , Q for disjunction (again, his word), meaning either P or Q . He then stated what are still known as *De Morgan’s laws*:

The contrary of PQ is p , q . *Not both* is either not one or not the other, or not either. *Not either P nor Q* (which we might denote by $: P , Q$ or $.P , Q$) is logically ‘*not P and not Q* ’ or pq : and this is then the contrary of P , Q .

45.2. SYMBOLIC CALCULUS

An example of the new freedom in the interpretation of symbols actually occurred somewhat earlier than the time of De Morgan, in Lagrange’s algebraic approach to analysis. Thinking of Taylor’s theorem as

$$\Delta_h f(x) = f(x + h) - f(x) = hDf(x)h + \frac{1}{2!}h^2 D^2 f(x) + \frac{1}{3!}h^3 D^3 f(x) + \dots ,$$

where $Df(x) = f'(x)$, and comparing with the Taylor series of the exponential function,

$$e^t = 1 + t + \frac{1}{2!}t^2 + \frac{1}{3!}t^3 + \dots ,$$

Lagrange arrived at the formal equation

$$\Delta_h = e^{hD} - 1 .$$

Although the equation is purely formal and should perhaps be thought of only as a convenient way of remembering Taylor’s theorem, it does suggest a converse relation

$$Df(x) = \frac{1}{h}(\ln(1 + \Delta_h))f(x) = \frac{1}{h}\left(\Delta_h f(x) + \frac{1}{2}\Delta_h^2 f(x) + \dots\right) ,$$

and this relation is literally true for polynomials $f(x)$. The formal use of this symbolic calculus may have been merely suggestive, but as Grattan-Guinness remarks (2000, p. 19), “some people regarded these methods as legitimate in themselves, not requiring foundations from elsewhere.”

45.3. BOOLE'S MATHEMATICAL ANALYSIS OF LOGIC

One such person was George Boole. In a frequently quoted passage from the introduction to his brief 1847 treatise *The Mathematical Analysis of Logic*, Boole wrote

[T]he validity of the processes of analysis does not depend upon the interpretation of the symbols which are employed but solely upon the laws of their combination. Every system of interpretation which does not affect the truth of the relations supposed is equally admissible, and it is thus that the same process may under one scheme of interpretation represent the solution of a question or the properties of number, under another that of a geometrical problem, and under a third that of optics.

Here Boole, like De Morgan, was arguing for the freedom to create abstract systems and attach an interpretation to them later. This step was still something of an innovation at the time. It was generally accepted, for example, that irrational and imaginary numbers had a meaning in geometry but not in arithmetic. One could not, or should not, simply *define* them into existence. Cayley raised this objection shortly after the appearance of Boole's treatise (see Grattan-Guinness, 2000, p. 41), asking whether it made any sense to write $\frac{1}{2}x$. Boole replied by comparing the question to the existence of $\sqrt{-1}$, which he said was "a symbol (*i*) which satisfies particular laws, and especially this: $i^2 = -1$." In other words, when we are inventing a formal system, we are nearly omnipotent. Whatever we prescribe will hold for the system we define. If we want a square root of -1 to exist, it will exist (whatever "exist" may mean).

45.3.1. Logic and Classes

Although set theory had different roots on the Continent, we can see its basic concept—membership in a class—in Boole's work. Departing from De Morgan's notation, he denoted a generic member of a class by an uppercase *X* and used the lowercase *x* "operating on any subject," as he said, to denote the class itself. Then xy was to denote the class "whose members are both *X*'s and *Y*'s." This language rather blurs the distinction between a set, its members, and the properties that determine what the members are; but we should expect that clarity would take some time to achieve. The connection between logic and set theory is an intimate one and one that is easy to explain. But the kind of set theory that logic alone would have generated was different from the geometric set theory of Georg Cantor, which was intimately connected with the topology of the real line.

The influence of the mathematical theory of probability on logic is both extensive and interesting. The subtitle of De Morgan's *Formal Logic* is *The Calculation of Inference, Necessary and Probable*, and, as noted above, three chapters (some 50 pages) of *Formal Logic* are devoted to probability and induction. Probability deals with events, whereas logic deals with propositions. The connection between the two was stated by Boole in his later treatise, *An Investigation of the Laws of Thought*, as follows:

[T]here is another form under which all questions in the theory of probabilities may be viewed; and this form consists in substituting for *events* the propositions which assert that those events have occurred, or will occur; and viewing the element of numerical probability as having reference to the *truth* of those *propositions*, not to the *occurrence* of the *events*.

Two events can combine in four different ways. Neither may occur, or E may occur but not F , or F may occur but not E , or E and F may both occur. If the events E and F are independent, the probability that both E and F occur is the *product* of their individual probabilities. If the two events cannot both occur, the probability that at least one occurs is the *sum* of their individual probabilities. More generally,

$$P(E \text{ or } F) + P(E \text{ and } F) = P(E) + P(F).$$

When these combinations of events are translated into logical terms, the result is a *logical calculus*.

The idea of probability 0 as indicating impossibility and probability 1 as indicating certainty must have had some influence on Boole's use of these symbols to denote "nothing" and "the universe." He expressed the proposition "all X 's are Y 's," for example, as $xy = x$ or $x(1 - y) = 0$. Notice that $1 - y$ appears, not $y - 1$, which would have made no sense. Here $1 - y$ corresponds to the things that are not- y . From there, it is not far to thinking of 0 as false and 1 as true. The difference between probability and logic here is that the probability of an event may be any number between 0 and 1, while propositions are either true or false.¹ These analogies were brought out fully in Boole's major work, to which we now turn.

45.4. BOOLE'S LAWS OF THOUGHT

Six years later, after much reflection on the symbolic logic that he and others had developed, Boole wrote an extended treatise, *An Investigation of the Laws of Thought*, which began by recapping what he had done earlier. The *Laws of Thought* began with a very general proposition that laid out the universe of symbols to be used. These were:

- 1st. Literal symbols, as x , y , &c., representing things as subjects of our conceptions.
- 2nd. Signs of operation, as $+$, $-$, \times , standing for those operations of the mind by which the conceptions of things are combined or resolved so as to form new conceptions involving the same elements.
- 3rd. The sign of identity, $=$.

And these symbols of Logic are in their use subject to definite laws, partly agreeing with and partly differing from the laws of the corresponding symbols in the science of Algebra.

Boole used $+$ to represent disjunction (or) and juxtaposition, used in algebra for multiplication, to represent conjunction (and). The sign $-$ was used to stand for "and not." In his examples, he used $+$ only when the properties were, as we would say, disjoint; and he

¹Classical set theory deals with propositions of the form $x \in E$, which are either true or false: Either x belongs to E , or it does not, and there is no other possibility. The recently created *fuzzy set theory* restores the analogy with probability, allowing an element to belong partially to a given class and expressing the degree of membership by a function $\varphi(x)$ whose values are between 0 and 1. Thus, for example, whether a woman is pregnant or not is a classical set-theory question; whether she is tall or not is a fuzzy set-theory question. Fuzzy-set theorists point out that their subject is not subsumed by probability, since it deals with the properties of individuals, not those of large sets.

used — only when the property subtracted was, as we would say, a subset of the property from which it was subtracted. He illustrated the equivalence of “European men and women” (where the adjective *European* is intended to apply to both nouns) with “European men and European women” as the equation $z(x + y) = zx + zy$. Similarly, to express the idea that the class of men who are non-Asiatic and white is the same as the class of white men who are not white Asiatic men, he wrote $z(x - y) = zx - zy$. He attached considerable importance to what he was later to call the *index law*, which expresses the fact that affirming a property twice conveys no more information than affirming it once. That is to say, $xx = x$, and he adopted the algebraic notation x^2 for xx . This piece of algebraization led him, by analogy with the rules $x0 = 0$ and $x1 = x$, to conclude that “the respective interpretations of the symbols 0 and 1 in the system of Logic are *Nothing* and *Universe*.” From these considerations he deduced the principle of contradiction: $x^2 = x \Rightarrow x(1 - x) = 0$, that is, no object can have a property and simultaneously not have that property.²

Boole was carried away by his algebraic analogies. Although he remained within the confines of his initial principles for a considerable distance, when he got to Chapter 5 he introduced the concept of *developing* a function. That is, for each algebraic expression $f(x)$, no matter how complicated, finding an equivalent linear expression $ax + b(1 - x)$, one that would have the same values as $f(x)$ for $x = 0$ and $x = 1$. That expression would obviously be $f(1)x + f(0)(1 - x)$. Boole gave a convoluted footnote to explain this simple fact by deriving it from Taylor’s theorem and the idempotence property.

45.5. JEVONS

Both De Morgan and Boole used the syllogism or *modus ponens* (*inferring method*) as the basis of logical inference, although De Morgan did warn against an overemphasis on it. William Stanley Jevons (1835–1882) formulated this law algebraically and adjoined to it a principle of indirect inference, which amounted to inference by exhaustive enumeration of cases. The possibility of doing the latter by sorting through slips of paper led him to the conclusion that this sorting could be done by machine. Since he had removed much of the mathematical notation used by Boole, he speculated that the mathematics could be entirely removed from it. He also took the additional step of suggesting, rather hesitantly, that mathematics was itself a branch of logic. According to Grattan-Guinness (2000, p. 59), this speculation apparently had no influence on the mathematical philosophers who ultimately developed its implications.

45.6. PHILOSOPHIES OF MATHEMATICS

Other mathematicians besides Cantor were also considering ways of deriving mathematics logically from simplest principles. Gottlob Frege (1848–1925), a professor in Jena, who occasionally lectured on logic, attempted to establish logic on the basis of “concepts” and

²Nowadays, a ring in which every element is idempotent—that is, the law $x^2 = x$ holds—is called a *Boolean ring*. It is an interesting exercise to show that such a ring is always commutative and of characteristic 2; that is, $x + x = 0$ for all x . The subsets of a given set form a Boolean ring when addition is interpreted as symmetric difference; that is, $A + B$ means “either A or B but not both.”

“relations” to which were attached the labels *true* or *false*. He was the first to establish a complete predicate calculus, and in 1884 wrote a treatise called *Grundgesetze der Arithmetik (Principles of Arithmetic)*. Meanwhile in Italy, Giuseppe Peano (1858–1939) was axiomatizing the natural numbers. Peano took the successor relation as fundamental and based his construction of the natural numbers on this one relation and nine axioms, together with a symbolic logic that he had developed. The work of Cantor, Frege, and Peano attracted the notice of a young student at Cambridge, Bertrand Russell (1872–1969), who had written his thesis on the philosophy of Leibniz. Russell saw in this work confirmation that mathematics is merely a prolongation of formal logic. This view, that mathematics can be deduced from logic without any new axioms or rules of inference, is now called *logicism*. Gödel’s work was partly a commentary on this program and the formalist program of Hilbert (discussed below) and can be interpreted as a counterargument to its basic thesis—that mathematics can be axiomatized. Logicism had encountered difficulties still earlier, however. Even the seemingly primitive notion of membership in a set turned out to require certain caveats.

45.6.1. Paradoxes

In 1897, Peano’s assistant Cesare Burali-Forti (1861–1931), apparently unintentionally, revealed a flaw in the ordinal numbers.³ To state the problem in the clear light of hindsight, if two ordinal numbers satisfy $x < y$, then $x \in y$, but $y \notin x$. In that case, what are we to make of the set of *all* ordinal numbers? Call this set A . Like any other ordinal number, it has a successor $A + 1$ and $A \in A + 1$. But since $A + 1$ is an ordinal number, we must also have $A + 1 \in A$, and hence $A < A + 1$ and $A + 1 < A$. This was the first paradox of uncritical set theory, but others were to follow.

The most famous paradox of set theory arose in connection with cardinal numbers rather than ordinal numbers. Cantor had defined equality between cardinal numbers as the existence of a one-to-one correspondence between sets representing the cardinal numbers. Set B has larger cardinality than set A if there is no function $f : A \rightarrow B$ that is “onto,” that is, such that every element of B is $f(x)$ for some $x \in A$. Cantor showed that the set of all subsets of A , which we denote 2^A , is always of larger cardinality than A , so that there can be no largest cardinal number. If $f : A \rightarrow 2^A$, the set $C = \{t \in A : t \notin f(t)\}$ is a subset of A , hence an element of 2^A , and it cannot be $f(x)$ for any $x \in A$. To see why, assume the opposite, that is, $C = f(x)$ for some x . Then either $x \in C$ or $x \notin C$. If $x \in C$, then $x \in f(x)$ and so by definition of C , $x \notin C$. On the other hand, if $x \notin C$, then $x \notin f(x)$, and again by definition of C , $x \in C$. Since the whole paradox results from the assumption that $C = f(x)$ for some x , it follows that no such x exists, that is, the mapping f is not “onto.” This argument was at first disputed by Russell, who wrote in an essay entitled “Recent work in the philosophy of mathematics” (1901) that “the master has been guilty of a very subtle fallacy.” Russell thought that there was a largest set, the set of *all* sets. In a later reprint of the article he added a footnote explaining that Cantor was right.⁴ Russell’s first attempt at a systematic exposition of mathematics as he thought it ought to be was

³Moore (1982, p. 59) notes that Burali-Forti himself did not see any paradox and (p. 53) that the difficulty was known earlier to Cantor.

⁴Moore (1982, p. 89) points out that Zermelo had discovered Russell’s paradox two years before Russell discovered it and had written to Hilbert about it. Zermelo, however, did not consider it a very troubling paradox. To him it meant only that no set should contain all of its subsets as elements.

his 1903 work *Principles of Mathematics*. According to Grattan-Guinness (2000, p. 311), Russell removed his objection to Cantor's proof and published his paradox in this work, but kept the manuscript of an earlier version, made before he was able to work out where the difficulty lay.

To explain Russell's mistake (as Russell later explained it to himself), consider the set of all sets. We must, by its definition, believe it to be *equal* to the set of all its subsets. Therefore the mapping $f(E) = E$ should have the property that Cantor says no mapping can have. Now if we apply Cantor's argument to this mapping, we are led to consider $S = \{E : E \notin E\}$. By definition of the mapping f we should have $f(S) = S$, and so, just as in the case of Cantor's argument, we ask if $S \in S$. Either way, we are led to a contradiction. This result is known as *Russell's paradox*.

After Russell had straightened out the paradox with a theory of types, he collaborated with his teacher Alfred North Whitehead (1861–1947) on a monumental derivation of mathematics from logic, published in 1910 as *Principia mathematica*.

45.6.2. Formalism

A different view of the foundations of mathematics was advanced by Hilbert, who was interested in the problem of axiomatization (the axiomatization of probability theory was the sixth of his famous 23 problems) and particularly interested in preserving as much as possible of the freedom to reason that Cantor had provided while avoiding paradoxes. The essence of this position, now known as formalism, is the idea stated by De Morgan and Boole that the legal manipulation of the symbols of mathematics and their interpretation are separate issues. Hilbert is famously quoted as having claimed that the words *point*, *line*, and *plane* should be replaceable by *table*, *chair*, and *beer mug* when a theorem is stated. Grattan-Guinness (2000, p. 208) notes that Hilbert may not have intended this statement in quite the way it is generally perceived and may not have thought the matter through at the time. He also notes (p. 471) that Hilbert never used the name *formalism*. Characteristic of the formalist view is the assumption that any mathematical object whatever may be defined, provided only that the definition does not lead to a contradiction. Cantor was a formalist in this sense (Grattan-Guinness, 2000, p. 119). In the formalist view, mathematics is the study of formal systems, but the rules governing those systems must be stated with some care. Formalism detaches the symbols and formulas of mathematics from the meanings attached to them in applications, making a distinction between syntax and semantics.

Hilbert had been interested in logical questions in the 1890s and early 1900s, but his work on formal languages such as propositional calculus dates from 1917. In 1922, when the intuitionists (discussed below) were publishing their criticism of mathematical methodologies, he formulated his own version of mathematical logic. In it, he introduced the concept of metamathematics, the study whose subject matter is the structure of a mathematical system.⁵ To avoid infinity in creating a formal language while preserving sufficient generality, Hilbert resorted to a "finitistic" device called a *schema*. Certain basic formulas are declared to be legitimate by fiat. Then a few rules are adopted, such as the rule that if A and B are legitimate formulas, so is $[A \Rightarrow B]$. This way of defining legitimate (well-formed) formulas makes it possible to determine in a finite number of steps whether or not a formula is well

⁵This distinction had been introduced by L. E. J. Brouwer in his 1907 thesis, but not given a name and never developed (see Grattan-Guinness, 2000, p. 481).

formed. It replaces the synthetic constructivist approach with an analytic approach (which can be reversed, once the analysis is finished, to synthesize a given well-formed formula from primitive elements).

The formalist approach makes a distinction between statements *of* arithmetic and statements *about* arithmetic. For example, the assertion that there are no positive integers x, y, z such that $x^3 + y^3 = z^3$ is a statement *of* arithmetic. The assertion that this statement can be proved from the axioms of arithmetic is a statement *about* arithmetic. The *metalanguage*, in which statements are made about arithmetic, contains all the meaning to be assigned to the propositions of arithmetic. In particular, it becomes possible to distinguish between what is true (that is, what can be known to be true from the metalanguage) and what is provable (what can be deduced within the object language). Two questions thus arise in the metalanguage: (1) *Is every deducible proposition true?* (the problem of consistency); (2) *Is every true proposition deducible?* (the problem of completeness). As we shall see below, Gödel showed that the answer, for first-order recursive arithmetic and more generally for systems of that type, is very pessimistic. This language is not complete and is incapable of proving its own consistency.

45.6.3. Intuitionism

The most cautious approach to the foundations of mathematics, known as *intuitionism*, was championed by the Dutch mathematician Luitzen Egbertus Jan Brouwer (1881–1966). Brouwer was one of the most mystical of mathematicians, and his mysticism crept into his early work. He even published a pamphlet in 1905, claiming that true happiness came from the inner world and that contact with the outer world brought pain (Franchella, 1995, p. 305). In his dissertation at the University of Amsterdam in 1907, he criticized the logicism of Russell and Zermelo's axiom of choice. Although he was willing to grant the validity of constructing each particular countable ordinal number, he questioned whether one could meaningfully form the set of *all* countable ordinals.⁶ In a series of articles published from 1918 to 1928, Brouwer laid down the principles of intuitionism. These principles include the rejection not only of the axiom of choice beyond the countable case, but also of proof by contradiction. That is, the implication "A implies not-(not-A)" is accepted, but not its converse, "Not-(not-A) implies A."

A specimen of Brouwer's philosophy may give some idea of the general trend of his thought. In lectures delivered at Cambridge University in 1946 (van Dalen, 1981), he defined a *fleeing property*⁷ to be a property f such that, for any given positive integer n it is possible to determine (via an algorithm) whether or not n has the property f , yet there is no known way of calculating any number possessing property f , and no known way of proving that no number has this property (van Dalen, 1981, pp. 6–7). As an example (not Brouwer's) let us take the property f for an integer n to mean that n is an odd perfect integer.

Brouwer recognized that the quality of being fleeing is not intrinsic to a property of the integers. It depends both on the definition of the property and on human history, and

⁶This objection seems strange at first, but the question of whether an effectively defined set must have effectively defined members is not at all trivial.

⁷Had English been his native language, Brouwer would probably have called this a *fugitive* property. In contrast to the English written by most of his compatriots, which is polished and elegant, Brouwer's was barbarous. He coined such atrocities as *noncontradictority* and *supposable*.

a property that is fleeing at one moment may cease to be fleeing at a later moment. For example, a positive integer that was a fifth power and simultaneously the sum of four other fifth powers was not known until the mid-twentieth century. These two conditions constituted a fleeing property until that time. They no longer do, since it is known that $27^5 + 84^5 + 110^5 + 133^5 = 144^5$. Proceeding from this definition, Brouwer defined the *critical* number κ_f for a fleeing property to be the smallest number having that property. (This is a number that is unknowable in principle as long as the property remains fleeing. If anyone ever produces even one example of a number with the property, it ceases to be fleeing.) He then further separates the positive integers into “up-numbers” of f , which are those at least as large as κ_f , and “down-numbers,” which are those that are smaller than κ_f . By the first part of the definition of a fleeing property, one can presumably establish that some initial segment of the positive integers consists of down-numbers. Brouwer then defines $a_\nu = 2^{-\nu}$ if ν is a down-number and $a_\nu = 2^{-\kappa_f}$ if ν is an up-number. Finally, he defines the number s_f to be the limit of a_ν as $\nu \rightarrow \infty$. According to Brouwer, this example refutes the principle that one of p or not- p must be true. For, he says,

...neither is $[s_f]$ equal to zero nor is it different from zero and, although its irrationality is absurd, it is not a rational number.

For a person accustomed to ordinary logic, this last statement is extraordinary. Brouwer appears to be conflating what is *known* with what is *true*, asserting, like the baseball umpire who says “They aren’t anything until I call them,” that this number does not have any properties until we know what those properties are. By ordinary logic, it is far from obvious that κ_f is defined. Indeed, in order to know what it is, we would have to annihilate it, since presumably κ_f is defined only while the property f is fleeing, and it ceases to be that once we know an integer that has property f . One is very much inclined to object that Brouwer has not really defined anything corresponding to the symbol κ_f , since the possibility exists that there are no numbers at all having property f , hence no smallest number with that property, and Brouwer has made no definition for κ_f in that case. In any case, the reader can see why intuitionism has not attracted a large number of adherents among mathematicians.

Intuitionists reject any proof whose implementation leaves choices to be made by the reader. Thus it is not enough in an intuitionist proof to say that objects of a certain kind exist. One must choose such an object and use it for the remainder of the proof. This extreme caution has rather drastic consequences. For example, the function $f(x)$ defined in ordinary language as

$$f(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

is not considered to be defined by the intuitionists, since there are ways of defining numbers x that do not make it possible to determine whether the number is negative or positive. For example, is the number $(-1)^n$, where n is the trillionth decimal digit of π , positive or negative? This restrictedness has certain advantages, however. The objects that are acceptable to the intuitionists tend to have pleasant properties. For example, every rational-valued function of a rational variable is continuous.

The intuitionist rejection of proof by contradiction needs to be looked at in more detail. Proof by contradiction was always used somewhat reluctantly by mathematicians, since such

proofs seldom give insight into the structures being studied. For example, the modern proof that there are infinitely many primes proceeds by assuming that the set of prime numbers is a finite set $P = \{p_1, p_2, \dots, p_n\}$ and showing that in this case the number $1 + p_1 \cdots p_n$ must either itself be a prime number or be divisible by a prime different from p_1, \dots, p_n , which contradicts the original assumption that p_1, \dots, p_n formed the entire set of prime numbers.

The appearance of starting with a false assumption and deriving a contradiction can be avoided here by stating the theorem as follows: If there exists a set of n primes p_1, \dots, p_n , there exists a set of $n + 1$ primes. The proof is exactly as before. Nevertheless, the proof is still not intuitionistically valid, since there is no way of saying whether or not $1 + p_1 \cdots p_n$ is prime. In intuitionistic logic, if “ p or q ” is a theorem, then either p is a theorem or q is a theorem.

In 1928 and 1929, a quarter-century after the debate over Zermelo’s axiom of choice, there was debate about intuitionism in the bulletin of the Belgian Royal Academy of Sciences. Two Belgian mathematicians, M. Barzin (dates unknown) and A. Errera (dates unknown), had argued that Brouwer’s logic amounted to a three-valued logic, since a statement could be true, false, or undecidable. The opposite point of view was defended by two Russian mathematicians, Aleksandr Yakovlevich Khinchin (1894–1959) and Valerii Ivanovich Glivenko (1897–1940). Barzin and Errera had suggested that to avoid three-valued logic, intuitionists ought to adopt as an axiom that if p implies “ q or r ,” then either p implies q or p implies r , and also that if “ p or q ” implies r , then p implies r and q implies r . Starting from these principles of Barzin and Errera and the trivial axiom “ p or not- p ” implies “ p or not- p ,” Khinchin deduced that p implies not- p and not- p implies p , thus reducing the suggestions of Barzin and Errera to nonsense. Glivenko took only a little longer to show that, in fact, Brouwer’s logic was not three-valued. He proved that the statement “ p or not- p is false” is false in Brouwer’s logic, and ultimately derived the theorem that the statement “ p is neither true nor false” is false (see Novosyolov, 2000).

A more “intuitive” objection to intuitionism is that intuition by its nature cannot be codified as a set of rules. In adopting such rules, the intuitionists were not being intuitionistic in the ordinary sense of the word. In any case, intuitionist mathematics is obviously going to be somewhat sparser in results than mathematics constructed on more liberal principles. That may be why it has attracted only a limited group of adherents.

45.6.4. Mathematical Practice

The paradoxes of naive set theory (such as Russell’s paradox) were found to be avoidable if the word *class* is used loosely, as Cantor had previously used the word *set*, but the word *set* is restricted to mean only a class that is a member of some other class. (Classes that are not sets are called *proper classes*.) Then to belong to a class A , a class B must not only fulfill the requirements of the definition of the class A but must also be known in advance to belong to some (possibly different) class.

This approach avoids Russell’s paradox. The class $C = \{x : x \notin x\}$ is a class; its elements are those classes that *belong to some class and* are not elements of themselves. If we now ask the question that led to Russell’s paradox—Is C a member of itself?—we do not reach a contradiction. If we assume $C \in C$, then we conclude that $C \notin C$, so that this assumption is not tenable. However, the opposite assumption, that $C \notin C$, is acceptable. It no longer leads to the conclusion that $C \in C$. For an object x to belong to C , it no longer suffices that $x \notin x$; it must also be true that $x \in A$ for some class A , an assumption not made for the case when x

is *C*. A complete set of axioms for set theory avoiding all known paradoxes was worked out by Paul Bernays (1888–1977) and Adolf Fraenkel (1891–1965). It forms part of the basic education of mathematicians today. It is generally accepted because mathematics can be deduced from it. However, it is very far from being a clear, concise, and therefore *obviously* consistent, foundation for mathematics. The axioms of set theory are extremely complicated and nonintuitive and are far less obvious than many things deduced from them. Moreover, their consistency is not only not obvious, it is even unprovable. In fact, one textbook of set theory, *Introduction to Set Theory*, by J. Donald Monk (McGraw-Hill, New York, 1969, p. 22), asserts regarding these axioms: “Naturally no inconsistency has been found, and *we have faith* that the axioms are, in fact, consistent”! (Emphasis added.)

45.7. DOUBTS ABOUT FORMALIZED MATHEMATICS: GÖDEL’S THEOREMS

The powerful and counterintuitive results obtained from the axiom of choice naturally led to doubts about the consistency of set theory. Since it was being inserted under the rest of mathematics as a foundation, the consistency question became an important one. A related question was that of completeness. Could one provide a foundation for mathematics, that is, a set of basic objects and rules of proof, that would allow any meaningful proposition to be proved true or false? The two desirable qualities of consistency and completeness are in the abstract opposed to each other, just as avoiding disasters and avoiding false alarms are opposing goals.

The most influential figure in mathematical logic during the twentieth century was Kurt Gödel (1906–1978). The problems connected with consistency and completeness of arithmetic, the axiom of choice, and many others all received a fully satisfying treatment at his hands that settled many old questions and opened up new areas of investigation. In 1931, he astounded the mathematical world by producing a proof that any consistent formal language in which arithmetic can be encoded is necessarily incomplete, that is, contains statements that are true according to its metalanguage but not deducible within the language itself. The intuitive idea behind the proof is a simple one, based on the following statement:

This statement cannot be proved.

Assuming that this statement has a meaning—that is, its context is properly restricted so that “proved” has a definite meaning—we can ask whether it is *true*. The answer must be positive if the system in which it is made is consistent. For if this statement is false, by its own content, it *can* be proved; and in a consistent deductive system, a false statement cannot be proved. Hence we agree that the statement is true, but, again by its own content, it cannot be proved.

The example just given is really nonsensical, since we have not stated the axioms and rules of inference that provide the context in which the statement is made. The word “proved” that it contains is not really defined. Gödel, however, took an accepted formalization of the axioms and rules of inference for arithmetic and showed that the metalanguage of arithmetic could be encoded within arithmetic. In particular each formula can be numbered uniquely, and the statement that formula n is (or is not) deducible from those rules can itself be coded as a well-formed formula of arithmetic. Then, when n is chosen so that the statement “Formula number n cannot be proved” happens to *be* formula n , we have exactly the situation just described. Gödel showed how to construct such an n . Thus, if Gödel’s version

of arithmetic is consistent, it contains statements that are formally undecidable. They are true (based on the metalanguage) but not deducible. This is Gödel's first incompleteness theorem. His second incompleteness theorem is even more interesting: *The assertion that arithmetic is consistent is one of the formally undecidable statements.*⁸ If the formalized version of arithmetic that Gödel considered is consistent, it is incapable of proving itself so. It is doubtful, however, that one could truly formalize every kind of argument that a rational person might produce. For that reason, care should be exercised in drawing inferences from Gödel's work to the actual practice of mathematics.

PROBLEMS AND QUESTIONS

Mathematical Problems

- 45.1. Suppose that the only allowable way of forming new formulas from old ones is to connect them by an implication sign; that is, given that A and B are well formed, $[A \Rightarrow B]$ is well formed, and conversely, if A and B are not both well formed, then neither is $[A \Rightarrow B]$. Suppose also that the only basic well-formed formulas are p , q , and r . Show that

$$[[p \Rightarrow r] \Rightarrow [p \Rightarrow q] \Rightarrow r]$$

is well formed but

$$[[p \Rightarrow r] \Rightarrow [r \Rightarrow]]$$

is not. Describe a general algorithm for determining whether a finite sequence of symbols is well formed.

- 45.2. Consider the following theorem. There exists an irrational number that becomes rational when raised to an irrational power. *Proof:* Consider the number $\theta = \sqrt{3}^{\sqrt{2}}$. If this number is rational, we have an example of such a number. If it is irrational, the equation $\theta^{\sqrt{2}} = \sqrt{3}^2 = 3$ provides an example of such a number. Is this proof intuitionistically valid?
- 45.3. Prove that $C = \{x : x \notin x\}$ is a proper class, not a set, that is, it is not an element of any class. (The assumption that it is an element of some class means it is a set, and then Russell's paradox results.)

Historical Questions

- 45.4. How did algebra influence the interaction of mathematics and logic starting in the nineteenth century?

⁸Detlefsen (2001) has analyzed the meaning of proving consistency in great detail and concluded that the generally held view of this theorem—that the consistency of a “sufficiently rich” theory cannot be proved by a “finitary” theory—is incorrect.

- 45.5. What was Charles Sanders Peirce's view of the relation between mathematics and logic?
- 45.6. What are the three major twentieth-century views of the philosophy of mathematics, and how do they differ from one another?

Questions for Reflection

- 45.7. Are there true but unknowable propositions in everyday life? Suppose that your class meets on Monday, Wednesday, and Friday. Suppose also that your instructor announces one Friday afternoon that you will be given a surprise exam at one of the regular class meetings the following week. One of the brighter students then reasons as follows. The exam will not be given on Friday, since if it were, having been told that it would be one of the three days, and not having had it on Monday or Wednesday, we would know on Thursday that it was to be given on Friday, and so it wouldn't be a surprise. Therefore it will be given on Monday or Wednesday. But then, since we *know* that it can't be given on Friday, it also can't be given on Wednesday. For if it were, we would know on Tuesday that it was to be given on Wednesday, and again it wouldn't be a surprise. Therefore it must be given on Monday, we know that now, and therefore it isn't a surprise. Hence it is impossible to give a surprise examination next week.

Obviously something is wrong with the student's reasoning, since the instructor can certainly give a surprise exam. Most students, when trying to explain what is wrong with the reasoning, are willing to accept the first step. That is, they grant that it is impossible to give a *surprise* exam on the *last* day of an assigned window of days. Yet they balk at drawing the conclusion that this argument implies that the originally next-to-last day must thereby become the last day. Notice that, if the professor had said nothing to the students, it would be possible to give a surprise exam on the last day of the window, since the students would have no way of knowing that there was any such window. The conclusion that the exam cannot be given on Friday therefore does not follow from assuming a surprise exam within a limited window alone, but rather from these assumptions supplemented by the following proposition: *The students know that the exam is to be a surprise and they know the window in which it is to be given.*

This fact is apparent if you examine the student's reasoning, which is full of statements about what the students *would know*. Can they truly *know* a statement (even a true statement) if it leads them to a contradiction?

Explain the paradox in your own words, deciding whether the exam would be a surprise if given on Friday. Can the paradox be avoided by saying that the conditions under which the exam is promised are true but the students cannot *know* that they are true?

How does this puzzle relate to Gödel's incompleteness result?

- 45.8. Brouwer, the leader of the intuitionist school of mathematicians, is also known for major theorems in topology, including the invariance of geometric dimension under homeomorphisms. One of his results is the *Brouwer fixed-point theorem*, which asserts that for any continuous mapping f of a closed disk into itself there is a point x such that $x = f(x)$. To prove this theorem, suppose there is a continuous mapping f for which $f(x) \neq x$ at every point x . Construct a continuous mapping g by drawing

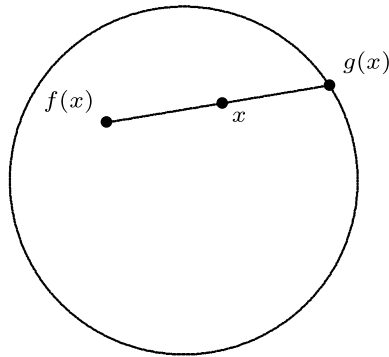


Figure 45.1. The Brouwer fixed-point theorem.

a line from $f(x)$ to x and extending it to the point $g(x)$ at which it meets the boundary circle (see Fig. 45.1). Then $g(x)$ maps the disk continuously onto its boundary circle and leaves each point of the boundary circle fixed. Such a continuous mapping is intuitively impossible (imagine stretching the entire head of a drum onto the rim without moving any point already on the rim and without tearing the head) and can be shown rigorously to be impossible (the disk and the circle have different homotopy groups). How can you explain the fact that the champion of intuitionism produced theorems that are not intuitionistically valid?

- 45.9.** Suppose that you prove a theorem by assuming that it is false and deriving a contradiction. What you have then proved is that either the axioms you started with are inconsistent or the assumption that the theorem is false is itself false. Why should you conclude the latter rather than the former? Is this why some mathematicians have claimed that the practice of mathematics requires faith?

Literature

Note: CSHPM/SCHPM = Canadian Society for the History and Philosophy of Mathematics/Société Canadienne d'Histoire et Philosophie des Mathématiques

- Adler, Ada, 1971. *Suidae Lexicon*, Teubner-Verlag, Stuttgart.
- Aḡargūin, Ahmet G.; Fletcher, Colin R., 1994. "Al-Farisi and the fundamental theorem of arithmetic," *Historia Mathematica*, **21**, No. 2, 162–173.
- Allman, George Johnston, 1889. *Greek Geometry from Thales to Euclid*, Longmans, Green & Co., London.
- Amir-Moez, Ali R., 1959. "Discussion of difficulties in Euclid by Omar ibn Abraham al-Khayyami," *Scripta Mathematica*, **XXIV**, 275–303.
- Amir-Moez, Ali R., 1963. "A paper of Omar Khayyam," *Scripta Mathematica*, **XXVI**, No. 4, 323–337.
- Andrews, George E., 1979. "An introduction to Ramanujan's 'lost' notebook," *American Mathematical Monthly*, **86**, No. 2, 89–108.
- Ang Tian-Se; Swetz, Frank J., 1986. "A Chinese mathematical classic of the third century: *The Sea Island Mathematical Manual* of Liu Hui," *Historia Mathematica*, **13**, No. 2, 99–117.
- Aschbacher, Michael, 1981. "The classification of the finite simple groups," *The Mathematical Intelligencer*, **3**, No. 2, 59–65.
- Ascher, Marcia, 1991. *Ethnomathematics*, Brooks/Cole, New York.
- Ascher, Marcia, 1997. "Malagasy *Sikidy*: A case in ethnomathematics," *Historia Mathematica*, **24**, No. 4, 376–395.
- Ayoub, R. 1980. "Paolo Ruffini's contributions to the quintic," *Archive for History of Exact Sciences*, **23**, No. 3, 253–277.
- Bag, Amulya Kumar, 1966. "Binomial theorem in ancient India," *Indian Journal of History of Science*, **1**, No. 1, 68–74.
- Bagheri, Mohammad, 1997. "A newly found letter of al-Kashi on scientific life in Samarkand," *Historia Mathematica*, **24**, No. 3, 241–256.
- Baigozhina, G. O., 1995. "On the classification principle of the problems in Abu-Kamil's *Book of Indeterminate Problems*," *Istoriko-Matematicheskie Issledovaniya*, **1**, No. 36, 61–66 (Russian).
- Baltzer, R., ed., 1885. *August Ferdinand Möbius, Gesammelte Werke*, S. Hirzel, Leipzig.

- Bashmakova, I. G; Smirnova, G. S., 1997. "The origin and development of algebra," in: B. V. Gnedenko, ed., *Essays on the History of Mathematics*, Moscow University Press, Moscow, pp. 94–246 (Russian). English translation published separately as *The Beginnings and Evolution of Algebra*, A. Shenitzer (transl.), Mathematical Association of America, Oberlin, OH, 2000.
- Beckers, Danny J., 1999. "Lagrange in the Netherlands: Dutch attempts to obtain rigor in calculus, 1797–1840," *Historia Mathematica*, **26**, No. 3, 234–238.
- Belcastro, Sarah-Marie; Yackel, Carolyn, 2008. *Making Mathematics with Needlework*, AK Peters, Wellesley, MA.
- Beman, Wooster Woodruff; Smith, David Eugene, 1930. *Famous Problems of Elementary Geometry*, G. E. Stechert & Co., New York.
- Berggren, J. L., 1986. *Episodes in the Mathematics of Medieval Islam*, Springer-Verlag, New York.
- Berggren, J. L., 1989. "Abu Sahl al-Kuhi: what the manuscripts say," *Proceedings of the 15th annual meeting of the CSHPM/SCHPM*, Université Laval, Montréal, Québec, pp. 31–48.
- Berggren, J. L., 1990. "Greek and Islamic elements in Arabic mathematics," *Proceedings of the 16th annual Meeting of the CSHPM/SCHPM*, University of Victoria, Victoria, British Columbia, pp. 25–38.
- Berggren, J. L., 2002. "The transmission of Greek geometry to medieval Islam," *CUBO*, **4**, No. 2, 1–13.
- Berggren, J. L.; Jones, Alexander, 2000. *Ptolemy's Geography*. Princeton University Press.
- Bernal, Martin, 1992. "Animadversions on the origins of western science," *Isis*, **83**, No. 4, 596–607.
- Bernays, Paul, ed., 1971. *Foundations of Geometry*, by David Hilbert, translated by Leo Unger, Open Court, La Salle, IL.
- Berndt, Bruce C., ed., 1985. *Ramanujan's Notebooks*, 5 vols., Springer-Verlag, New York.
- Betti, E. 1852. "Sulla risoluzione delle equazioni algebriche," *Tortolini Annali*, **III**, 49–51.
- Blackwell, Richard, transl., 1986. *Christiaan Huygens' The Pendulum Clock*, Iowa State University Press, Ames, IA.
- Bochenski, I. M., 1961. *A History of Formal Logic*, University of Notre Dame Press.
- Boncompagni, Baldassare, 1854. *Intorno ad alcune opere matematica notizie di Leonardo, matematico del secolo decimoterzo*, Tipografia Delle Belle Arti, Rome.
- Bottazzini, Umberto, 1986. *The Higher Calculus: A History of Real and Complex Analysis from Euler to Weierstrass*, Springer-Verlag, New York.
- Boyer, Carl B., 1949. *The History of the Calculus and its Conceptual Development*, Hafner, New York. Reprint: Dover, New York, 1959.
- Brentjes, Sonja; Hogendijk, Jan P., 1989. "Notes on Thabit ibn Qurra and his rule for amicable numbers," *Historia Mathematica*, **16**, No. 4, 373–378.
- Bretschneider, Carl Anton, 1870. *Die Geometrie und die Geometer vor Euklides. Ein historischer Versuch*, Teubner, Leipzig. Reprint: M. Sändig, Wiesbaden, 1968.
- Brett, William F.; Feldman, Emile B.; Sentlowitz, Michael, 1974. *An Introduction to the History of Mathematics, Number Theory and Operations Research*, MSS Information Corporation, New York.
- Brown, James Robert, 1999. *Philosophy of Mathematics: An Introduction to the World of Proofs and Pictures*, Routledge, New York.
- Buck, R. C., 1980. "Sherlock Holmes in Babylon," *The American Mathematical Monthly*, **87**, No. 5, 335–345.
- Burington, Richard Stevens, 1958. *Handbook of Mathematical Tables and Formulas*, Handbook Publishers, Sandusky, OH.

- Burkert, Walter, 1962. *Weisheit und Wissenschaft: Studien zu Pythagoras, Philolaos und Platon*. Verlag Hans Carl, Nürnberg.
- Butzmann, Hans, 1970. *Codex Agrimensorum Romanorum: Codex Arcerianusa der Herzog-August-Bibliothek zu Wolfenbüttel*, A. W. Sijthoff, Lugduni Batavorum.
- Bychkov, S. N., 2001. "Egyptian geometry and Greek science," *Istoriko-Matematicheskie Issledovaniya*, **6**, No. 41, 277–284 (Russian).
- Cajori, Florian, 1922. *A History of Mathematics*. Macmillan & Company, New York.
- Cantor, Moritz, 1880. *Vorlesungen über Geschichte der Mathematik*, Vol. 1, Teubner, Leipzig.
- Cauchy, A.-L., 1815. "Mémoire sur le nombre des valeurs qu'une fonction peut acquérir," *Journal de l'École Polytechnique, XVII^e Cahier, Tome X = Œuvres Complètes, Tome 13*, pp. 64–90. Gauthier-Villars, Paris.
- Caveing, Maurice, 1985. "La tablette babylonienne AO 17264 du Musée du Louvre et le problème des six frères," *Historia Mathematica*, **12**, No. 1, 6–24.
- Chace, A. B., Bull, L., Manning, H. P., and Archibald, R. C., 1927. *The Rhind Mathematical Papyrus*, Mathematical Association of America, Oberlin, OH, Vol. 1.
- Chaikovskii, Yu. V., 2001. "What is probability? The evolution of the concept (from antiquity to Poisson)," *Istoriko-Matematicheskie Issledovaniya*, **6**, No. 41, 34–56 (Russian).
- Chaucer, Geoffrey, 1391. *A Treatise on the Astrolabe*, in: F. N. Robinson, ed., *The Works of Geoffrey Chaucer*, 2nd ed., Houghton Mifflin, Boston, 1957, pp. 545–563.
- Chemla, Karine, 1991. "Theoretical aspects of the Chinese algorithmic tradition (first to third century)," *Historia Scientiarum*, No. 42, 75–98.
- Christianidis, Jean, 1998. "Une interprétation byzantine de Diophante," *Historia Mathematica*, **25**, No. 1, 22–28.
- Ciancio, Salvatore, 1965. *La tomba di Archimede*, Ciranna, Roma.
- Clagett, Marshall, 1960. *The Medieval Science of Weights*, University of Wisconsin Press, Madison, WI.
- Clagett, Marshall, 1968. *Nicole Oresme and the Medieval Geometry of Qualities and Motions*, University of Wisconsin Press, Madison, WI.
- Clark, Walter Eugene, ed., 1930. *The Aryabhatiya of Aryabhata*, University of Chicago Press, Chicago.
- Closs, Michael P., ed., 1986. *Native American Mathematics*, University of Texas Press, Austin, TX.
- Closs, Michael P., 1992. "Ancient Maya mathematics and mathematicians," *Proceedings of the 18th Annual Meeting of the CSHPM/SCHPM*, University of Prince Edward Island, Charlottetown, P.E.I., pp. 1–13.
- Coe, Michael D., 1973. *The Maya Scribe and His World*, The Grolier Club, New York.
- Colebrooke, Henry Thomas, 1817. *Algebra with Arithmetic and Mensuration from the Sanscrit of Brahme Gupta and Bhascara*, J. Murray, London.
- Colson, F. H., 1926. *The Week: An Essay on the Origin and Development of the Seven-Day Cycle*, Greenwood Press, Westport, CT.
- Coolidge, Julian Lowell, 1940. *A History of Geometrical Methods*, Clarendon Press, Oxford.
- Craik, Alex D. D., 1999. "Calculus and analysis in early nineteenth-century Britain: the work of William Wallace," *Historia Mathematica*, **26**, No. 3, 239–267.
- Crossley, John N.; Henry, Alan S., 1990. "Thus spake al-Khwarizmi: A translation of the text of Cambridge University Library Ms. ii.vi.5," *Historia Mathematica*, **17**, No. 2, 103–131.
- Cullen, Christopher, 1996. *Astronomy and Mathematics in Ancient China: The Zhou Bi Suan Jing*, Cambridge University Press.

- Cuomo, Serafina, 2000. *Pappus of Alexandria and the Mathematics of Late Antiquity*, Cambridge University Press.
- al-Daffa, Ali Abdullah, 1977. *The Muslim Contribution to Mathematics*, Humanities Press, Atlantic Highlands, NJ.
- Dahan, Amy, 1980. "Les travaux de Cauchy sur les substitutions; Étude de son approche du concept de groupe," *Archive for History of Exact Sciences*, **23**, No. 4, 279–319.
- Dahan-Dalmédico, Amy, 1987. "Mécanique et théorie des surfaces: les travaux de Sophie Germain," *Historia Mathematica*, **14**, No. 4, 347–365.
- van Dalen, D., 1981. *Brouwer's Cambridge Lectures on Intuitionism*, Cambridge University Press.
- Dauben, Joseph W., 1996. "Mathematics at the University of Toronto: Abraham Robinson in Canada (1951–1957)," in: Dauben, Folkerts, Knobloch, and Wussing, *History of Mathematics: States of the Art*, Academic Press, New York.
- Dauben, Joseph W.; Scriba, Christoph J., eds., 2002. *Writing the History of Mathematics: Its Historical Development*, Birkhäuser, Boston.
- David, H. A.; Edwards, A. W. F., 2001. *Annotated Readings in the History of Statistics*, Springer-Verlag, New York.
- Davis, Margaret Daly, 1977. *Piero Della Francesca's Mathematical Treatises: the Tratto d'abaco and Libellus de quinque corporibus regularibus*, Longe Editore, Ravenna.
- Davis, Philip J.; Hersh Reuben, 1986. *Descartes' Dream: The World According to Mathematics*, Harcourt Brace Jovanovich, New York.
- Deakin, Michael, 1994. "Hypatia and her mathematics," *American Mathematical Monthly*, **101**, No. 3, 234–243.
- Detlefsen, Michael, 2001. "What does Gödel's second theorem say?" *Philosophia Mathematica* (3), **9**, No. 1, 37–71.
- Detlefsen, Michael; Erlandson, Douglas K.; Heston, J. Clark; Young, Charles M., 1975. "Computation with Roman numerals," *Archive for History of Exact Science*, **15**, No. 2, 141–148.
- Devlin, Keith, 2011. *The Man of Numbers: Fibonacci's Arithmetic Revolution*, Walker and Company, New York.
- De Mora-Charles, 1992. "Quelques jeux de hazard selon Leibniz," *Historia Mathematica* **19**, No. 2, 125–157.
- de Mora-Charles, S., 1992. "Quelques jeux de hazard selon Leibniz," *Historia Mathematica*, **19**, No. 2, 125–157.
- De Young, Gregg, 1995. "Euclidean geometry in the tradition of Islamic India," *Historia Mathematica*, **22**, No. 2, 138–153.
- Dick, Auguste, 1981. *Emmy Noether, 1882–1935*, translated by H. I. Blocher. Birkhäuser, Boston.
- Dickson, Leonard Eugene, 1919. *History of the Theory of Numbers I: Divisibility and Primality*, Carnegie Institute, Washington, DC. Reprint: Chelsea, New York, 1966.
- Dickson, Leonard Eugene, 1920. *History of the Theory of Numbers II: Diophantine Analysis*, Carnegie Institute, Washington, DC. Reprint: Chelsea, New York, 1966.
- Dickson, Leonard Eugene, 1923. *History of the Theory of Numbers III: Quadratic and Higher Forms*, Carnegie Institute, Washington, DC. Reprint: Chelsea, New York, 1966.
- Diels, Hermann, 1951. *Die Fragmente der Vorsokratiker*, 6th corrected edition (Walther Kranz, ed.), Weidmann, Berlin.
- Dijksterhuis, E. J., 1956. *Archimedes*, Munksgård, Copenhagen.
- Dilke, O. A. W., 1985. *Greek and Roman Maps*, Cornell University Press, Ithaca, NY.
- D'ooze, Martin Luther, 1926, translator. *Introduction to Arithmetic* (Nicomachus of Gerasa), Macmillan, New York.

- Dorofeeva, A. V., 1998. "The calculus of variations," in: *Mathematics of the 19th Century*, Birkhäuser, Basel, pp. 197–260.
- Duren, Peter, 1989. *A Century of Mathematics in America* (3 Vols.), American Mathematical Society, Providence, RI.
- Dutka, Jacques, 1988. "On the Gregorian revision of the Julian calendar," *Mathematical Intelligencer*, **10**, No. 1, 56–64.
- Dzielska, Maria, 1995. *Hypatia of Alexandria*, Harvard University Press, Cambridge, MA.
- Edwards, H. M., 1974. *Riemann's Zeta Function*, Academic Press, New York.
- Edwards, H. M., 1977. *Fermat's Last Theorem*, Springer-Verlag, New York.
- Engel, Friedrich; Heegaard, Poul, eds., 1960. *Sophus Lie, Gesammelte Abhandlungen*, Teubner, Leipzig.
- Erdős, P.; Dudley, U., 1983. "Some remarks and problems in number theory related to the work of Euler," *Mathematics Magazine*, **56**, No. 5, 292–298.
- Euler, L., 1732. "De formes radicum aequationum cuiusque ordinis coniectatio," *Commentarii Academiae Petropolitanae*, **6** (1738), p. 216.
- Euler, L., 1744. *Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes*, Bousquet & Co., Lausanne.
- Euler, L., 1749. "Recherches sur les racines imaginaires des équations," *Histoire de l'Académie des Sciences de Berlin*, p. 222.
- Euler, L., 1762. "De resolutione aequationum cuiusque gradus," *Novi Commentarii Academiae Petropolitanae*, **9**, p. 70.
- Feigenbaum, L., 1994. "Infinite series and solutions of ordinary differential equations, 1670–1770," in: *Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences*, Vol. 1, Routledge, London and New York, pp. 504–519.
- Feingold, Mordechai, 1993. "Newton, Leibniz, and Barrow too: An attempt at a reinterpretation," *Isis*, **84**, No. 2, 310–338.
- Ferreirós, José, 1995. "'What fermented in me for years': Cantor's discovery of transfinite numbers," *Historia Mathematica*, **22**, 33–42.
- Field, J. V.; Gray, J. J., 1987. *The Geometrical Work of Girard Desargues*, Springer-Verlag, Berlin.
- Fitzgerald, Augustine, transl., 1926. *The Letters of Synesius of Cyrene*, Oxford University Press.
- Flegg, G. 1988. "Nicolas Chuquet—An introduction," in: Cynthia Hay, ed., *Mathematics from Manuscript to Print, 1300–1600*, Oxford University Press, pp. 59–72.
- Fletcher, Colin R., 1989. "Fermat's theorem," *Historia Mathematica*, **16**, No. 2, 149–153.
- Folkerts, Menso, 1970. *Ein neuer Text des Euclides latinus*, H. A. Gerstenberg, Hildesheim.
- Folkerts, Menso, 1971. *Anonyme lateinische Euklidbearbeitungen aus dem 12. Jahrhundert*, Österreichische Akademie der Wissenschaften, Mathematisch-Naturwissenschaftliche Klasse, Denkschriften, **116**, erste Abhandlung.
- Fowler, David, 1992. "Dedekind's theorem: $\sqrt{2} \times \sqrt{3} = \sqrt{6}$," *The American Mathematical Monthly*, **99**, No. 8, 725–733.
- Fowler, David, 1998. *The Mathematics of Plato's Academy*, 2nd ed., Clarendon Press, Oxford.
- Franchella, Miriam, 1995. "L. E. J. Brouwer: toward intuitionistic logic," *Historia Mathematica*, **22**, No. 3, 304–322.
- Franci, Rafaella, 1988. "Antonio de' Mazzinghi, an algebraist of the fourteenth century," *Historia Mathematica*, **15**, No. 3, 240–249.
- Fraser, Craig, 1987. "Joseph-Louis Lagrange's algebraic vision of the calculus," *Historia Mathematica*, **14**, No. 1, 38–53.

- Fraser, Craig, 1993. "A history of Jacobi's theorem in the calculus of variations," *Proceedings of the 19th Annual Meeting of the CSPHM/SCHPM*, Carleton University, Ottawa, Ontario, pp. 168–185.
- Friberg, Jöran, 1981. "Methods and traditions of Babylonian mathematics: Plimpton 322, Pythagorean triples and the Babylonian triangle parameter equations," *Historia Mathematica*, **8**, No. 3, 277–318.
- Fried, Michael N.; Unguru, Sabetai, 2001. *Apollonius of Perga's Conica. Text, Context, Subtext*, Brill, Boston.
- von Fritz, Kurt, 1945. "The discovery of incommensurability by Hippasus of Metapontum," *Annals of Mathematics*, **46**, 242–264.
- Fu Daiwie, 1991. "Why did Lui Hui fail to derive the volume of a sphere?," *Historia Mathematica*, **18**, No. 3, 212–238.
- Fukugawa Hidetoshi, ed., 2005. *Sho Min No San Jutsu Ten (Exhibit of Popular Computational Art)*, Nagoya City Museum of Science.
- Fukagawa Hidetoshi; Pedoe, D., 1989. *Japanese Temple Geometry Problems: San Gaku*, Winnipeg, Manitoba.
- Fukagawa Hidetoshi; Rothman, T. 2008. *Sacred Mathematics: Japanese Temple Geometry*, Princeton University Press. *Sho Min No San Jutsu Ten (Exhibit of Popular Computational Art)*
- Fuson, Karen, 1988. *Children's Counting and Concepts of Number*, Springer-Verlag, New York.
- Gandz, Solomon, 1926. "The origin of the term 'algebra'," *American Mathematical Monthly*, **33**, No. 9, 437–440.
- Gauss, Carl Friedrich, 1799. "Demonstratio nova theorematis omnem functionem algebraicam rationalem integram unius variabilis in factores reales primi vel secundi gradus resolvi posse," in: *Werke*, Vol. 3, Königlichen Gesellschaft der Wissenschaften, Göttingen, 1866, pp. 3–31.
- Gauss, Carl Friedrich, 1965. *General Investigations of Curved Surfaces*, translated from the Latin and German by Adam Hildebrandt and James Morehead, Raven Press, Hewlett, NY.
- Geijsbeek, John B., ed. and transl., 1914. *Ancient Double-Entry Bookkeeping. Lucas Pacioli's Treatise (A.D. 1494—The Earliest Known Writer on Bookkeeping)*, Denver, CO.
- Gerdes, Paulus, 1985. "Three alternate methods of obtaining the ancient Egyptian formula for the area of a circle," *Historia Mathematica*, **12**, No. 3, 261–268.
- Gerhardt, C. J., ed., 1971. *Leibniz: Mathematische Schriften*, G. Olms Verlag, Hildesheim.
- Gericke, Helmut, 1996. "Zur Geschichte der negativen Zahlen," in: Dauben, Folkerts, Knobloch, and Wussing, eds., *History of Mathematics: States of the Art*, Academic Press, New York, pp. 279–306.
- Gericke, Helmut; Vogel, Kurt, 1965. *De Thiende von Simon Stevin*, Akademische Verlagsgesellschaft, Frankfurt am Main.
- Gillings, Richard J., 1972. *Mathematics in the Time of the Pharaohs*, MIT Press, Cambridge, MA. Reprint: Dover, New York, 1982.
- Gold, David; Pingree, David, 1991. "A hitherto unknown Sanskrit work concerning Mādhava's derivation of the power series for sine and cosine," *Historia Scientiarum*, No. 42, 49–65.
- Goldstine, Herman H., 1980. *A History of the Calculus of Variations from the 17th Through the 19th Century*, Springer-Verlag, New York.
- Gottwald, Siegfried; Ilgauds, Hans-Joachim; Schlote, Karl-Heinz, eds., 1990. *Lexikon Bedeutender Mathematiker*, Bibliographisches Institut, Leipzig.
- Gould, James L.; Gould, Carol Grant, 1995. *The Honey Bee*, Scientific American Library, New York.
- Gow, James, 1884. *A Short History of Greek Mathematics*. Reprint: Chelsea, New York, 1968.
- Grabiner, Judith, 1981. *The Origins of Cauchy's Rigorous Calculus*. MIT Press.
- Grabiner, Judith, 1995. "Descartes and problem-solving," *Mathematics Magazine*, **68**, No. 2, 83–97.
- Grattan-Guinness, Ivor, 1972. "A mathematical union: William Henry and Grace Chisholm Young," *Annals of Science*, **29**, No. 2, 105–186.

- Grattan-Guinness, Ivor, 1975. "Mathematical bibliography for W. H. and G. C. Young," *Historia Mathematica*, **2**, 43–58.
- Grattan-Guinness, Ivor, 1990. *Convolutions in French Mathematics, 1800–1840*, Birkhäuser, Basel.
- Grattan-Guinness, Ivor, 2000. *The Search for Mathematical Roots, 1870–1940: Logics, Set Theories, and the Foundations of Mathematics from Cantor through Russell to Gödel*, Princeton University Press, Princeton, NJ.
- Grattan-Guinness, Ivor, 2004. "History or heritage? An important distinction in mathematics and for mathematics education," *American Mathematical Monthly*, **111**, 1–12.
- Gray, J. J., 1989. *Ideas of Space: Euclidean, Non-Euclidean, and Relativistic*, 2nd ed., Clarendon Press, Oxford.
- Gray, J. J., 2005. "Bernhard Riemann, posthumous thesis 'On the hypotheses which lie at the foundations of geometry'," in: I. Grattan-Guinness, ed., *Landmark Writings in Western Mathematics, 1640–1940*, Elsevier, Amsterdam, 2005, pp. 506–520.
- Gray, Robert, 1994. "Georg Cantor and transcendental numbers," *The American Mathematical Monthly*, **101**, No. 9, 819–832.
- Greenleaf, Benjamin, 1876. *New Practical Arithmetic; in Which the Science and Its Applications Are Simplified by Induction and Analysis*, Leach, Shewell, and Sanborn, Boston and New York.
- Grosholz, Emily, 1987. "Two Leibnizian manuscripts of 1690 concerning differential equations," *Historia Mathematica*, **14**, No. 1, 1–37.
- Guizal, Brahim; Dudley, John, 2002. "Ibn Sahl: Inventeur de la loi de la réfraction," *Revue pour la science*, No. 301, November 2002.
- Guo Shuchung, 1992. "Guo Shuchung's edition of the *Jiu Zhang Suan Shu*," *Historia Mathematica*, **19**, No. 2, 200–202.
- Gupta, R. C., 1989. "Sino-Indian interaction and the great Chinese Buddhist astronomer-mathematician I-Hsing (A.D. 683–727)," *Bulletin of the Indian Society for History of Mathematics*, **11**, Nos. 1–4, 38–49.
- Gupta, R. C., 1991. "On the volume of a sphere in ancient India," *Historia Scientiarum*, No. 42, 33–44.
- Gupta, R. C., 1994a. "Six types of Vedic mathematics," *Bulletin of the Indian Society for History of Mathematics*, **16**, Nos. 1–4, 5–15.
- Gupta, R. C., 1994b. "A circulatory rule from the *Agni Purāṇa*," *Bulletin of the Indian Society for History of Mathematics*, **16**, Nos. 1–4, 53–56.
- Gustafson, W. H.; Halmos, P. R.; Moolgavkar, S. H.; Wheeler, W. H.; Ziemer, W. P., 1976. "American mathematics from 1940 to the day before yesterday," *The American Mathematical Monthly*, **83**, No. 7, 503–516.
- Hadamard, J., 1896. "Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques," *Bulletin de la Société Mathématique de France*, **24**, 199–220.
- Hairetdinova, N. G., 1986. "On spherical trigonometry in the medieval Near East and in Europe," *Historia Mathematica*, **13**, No. 2, 136–146.
- Hamilton, W. R., 1837. "On the argument of Abel, respecting the impossibility of expressing a root of any general equation above the fourth degree, by any finite combination of radicals and rational functions," *Transactions of the Royal Irish Academy*, **XVIII** (1839), 171–259; H. Halberstam and R. Ingram, eds., *The Mathematical Papers of Sir William Rowan Hamilton*, Vol. III, pp. 517–569.
- Hari, K. Chandra, 2002. "Genesis and antecedents of Āryabhaṭīya," *Indian Journal of History of Science*, **37**, No. 2, 101–113.
- al-Hassan, Ahmad Y.; Hill, Donald R., 1986. *Islamic Technology: An Illustrated History*, Cambridge University Press.

- Hawkins, Thomas, 1989. "Line geometry, differential equations, and the birth of Lie's theory of groups," in: David Rowe and John McCleary, eds., *The History of Modern Mathematics*, Vol. 1, Academic Press, New York.
- Hayashi Takao, 1987. "Varahamihira's pandiagonal magic square of the order four," *Historia Mathematica*, **14**, No. 2, 159–166.
- Hayashi Takao, 1991. "A note on Bhāskara I's rational approximation to sine," *Historia Scientiarum*, No. 42, pp. 45–48.
- Heath, T.L., 1910. *Diophantus of Alexandria: A Study in the History of Greek Algebra*, 2nd ed., Cambridge University Press, 1910.
- Heath, T.L., 1897–1912. *The Works of Archimedes Edited in Modern Notation with Introductory Chapters*, Reprint: Dover, New York, 1953.
- Heath, T.L., 1921. *A History of Greek Mathematics*, Clarendon Press, Oxford.
- Hersh, Reuben, 1997. *What is Mathematics, Really?* Oxford University Press, New York.
- Heyde, C. C.; Seneta, E., 1977. *I. J. Bienaymé: Statistical Theory Anticipated*, Springer-Verlag, New York.
- Hilbert, David, 1971. *Foundations of Geometry*, translated by Leo Unger, Open Court, LaSalle, IL.
- Hille, Einar, 1959. *Analytic Function Theory*, Vol. 1. Ginn and Company, New York.
- Hogendijk, Jan P., 1985. "Thabit ibn Qurra and the pair of amicable numbers 17296, 18416," *Historia Mathematica*, **12**, No. 3, 269–273.
- Hogendijk, Jan P., 1989. "Sharaf al-Din al-Tusi on the number of positive roots of cubic equations," *Historia Mathematica*, **16**, No. 1, 69–85.
- Hogendijk, Jan P., 1991. "Al-Khwarizmi's table of the 'sine of the hours' and underlying sine table," *Historia Scientiarum*, No. 42, pp. 1–12.
- Hogendijk, Jan P., 2002. "The surface area of the bicylinder and Archimedes' *Method*," *Historia Mathematica*, **29**, No. 1, 199–203.
- Holt, P. M.; Lambton, Ann K. S.; Lewis, Bernard, eds., 1970. *The Cambridge History of Islam*, Vol. 2B.
- Homann, Frederick A., 1987. "David Rittenhouse: logarithms and leisure," *Mathematics Magazine*, **60**, No. 1, 15–20.
- Homann, Frederick A., 1991. *Practical Geometry: Practica Geometriae, attributed to Hugh of St. Victor*, Marquette University Press, Milwaukee, WI.
- Høyrup, Jens, 2002. "A note on Old Babylonian computational techniques," *Historia Mathematica*, **29**, No. 2, 193–198.
- Høyrup, Jens, 2010. "Old Babylonian 'Algebra,' and What It Teaches Us about Possible Kinds of Mathematics," Paper presented at the ICM Satellite Conference on Mathematics in Ancient Times, Kerala School of Mathematics, 19 August–1 September 2010. Available on-line at www.akira.ruc.dk/~jensh.
- Hughes, Barnabas, 1981. *De numeris datis* (Jordan de Nemore), University of California Press, Berkeley, CA.
- Hughes, Barnabas, 1989. "The arithmetical triangle of Jordanus de Nemore," *Historia Mathematica*, **16**, No. 3, 213–223.
- Hultsch, F., ed., 1965. *Pappi Alexandrini Collectionis*, Vol. 1, Verlag Adolf M. Hakkert, Amsterdam.
- Huygens, Christiaan, 1888. *Œuvres complètes*. M. Nijhoff, La Haye.
- Ifrah, Georges, 2000. *The Universal History of Numbers: From Prehistory to the Invention of the Computer*, translated from the French by David Bellos, Wiley, New York.

- Il'ina, E. A., 2002. "On Euclid's *Data*," *Istoriko-Matematicheskie Issledovaniya*, **7**, No. 42, 201–208 (Russian).
- Indorato, Luigi; Nastasi, Pietro, 1989. "The 1740 resolution of the Fermat–Descartes controversy," *Historia Mathematica*, **16**, No. 2, 137–148.
- Ivić, Aleksandar, 1985. *The Riemann Zeta-Function: Theory and Applications*, Dover, New York.
- Ivins, W. M., 1947. "A note on Desargues' theorem," *Scripta Mathematica*, **13**, 203–210.
- Jami, Catherine, 1988. "Western influence and Chinese tradition in an eighteenth century Chinese mathematical work," *Historia Mathematica*, **15**, No. 4, 311–331.
- Jami, Catherine, 1991. "Scholars and mathematical knowledge during the late Ming and early Qing," *Historia Scientiarum*, No. 42, 95–110.
- Jentsch, Werner, 1986. "Auszüge aus einer unveröffentlichten Korrespondenz von Emmy Noether und Hermann Weyl mit Heinrich Brandt," *Historia Mathematica*, **13**, No. 1, 5–12.
- Jha, V. N., 1994. "Indeterminate analysis in the context of the Mahāsiddhānta of Āryabhaṭa II," *Indian Journal of History of Science*, **29**, No. 4, 565–578.
- Jones, Alexander, 1986. *Book 7 of the Collection*, Springer-Verlag, New York.
- Jones, Alexander, 1990. *Ptolemy's First Commentator*, American Philosophical Society, Philadelphia.
- Jones, Alexander, 1991. "The adaptation of Babylonian methods in Greek numerical astronomy," *Isis*, **82**, No. 313, 441–453.
- Kasir, Daoud, 1931. *The Algebra of Omar Khayyam*, Teachers College, Columbia University Contributions to Education, No. 385, New York.
- Kawahara Hideki, 1991. "World-View of the *Santong-Li*," *Historia Scientiarum*, No. 42, 67–73.
- Kazdan, Jerry L., 1986. "A visit to China," *The Mathematical Intelligencer*, **8**, No. 4, 22–32.
- Kimberling, C. H., 1972a. "Emmy Noether," *The American Mathematical Monthly*, **79**, No. 2, 136–149.
- Kimberling, C. H., 1972b. "Addendum to 'Emmy Noether'," *The American Mathematical Monthly*, **79**, No. 7, 755.
- King, R. Bruce, 1996. *Beyond the Quartic Equation*, Birkhäuser, Boston.
- Kiro, S. N., 1967, "N. I. Lobachevskii and mathematics at Kazan' University," in: *History of Russian and Soviet Mathematics (Istoriya Otechestvennoi Matematiki)*, Vol. 2, Naukova Dumka, Kiev (Russian).
- Klein, Felix, 1884. *Lectures on the Icosahedron and the Solution of Equations of the Fifth Degree*, translated by George Gavin Morrice. Reprint: Dover, New York, 1956.
- Klein, Felix, 1897. *The Theory of the Top*. Charles Scribner's Sons, New York.
- Klein, Felix, 1926. *Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert*, Springer-Verlag, Berlin, 2 vols. Reprint: American Mathematical Society (Chelsea Publishing Company), Providence, RI, 1967.
- Klein, Jacob, 1933. *Plato's Trilogy: Theaetetus, the Sophist, and the Statesman*, University of Chicago Press, Chicago, 1977.
- Klein, Jacob, 1934–1936. *Greek Mathematical Thought and the Origin of Algebra*, translated by Eva Brann, MIT Press, Cambridge, MA, 1968.
- Klein, Jacob, 1965. *A Commentary on Plato's Meno*, University of North Carolina Press, Chapel Hill, NC.
- Kleiner, Israel, 1991. "Emmy Noether: Highlights of her life and work," *Proceedings of the 17th Annual Meeting of the CSHPM/SCHPM*, Queen's University, Kingston, Ontario, pp. 19–42.
- Kline, Morris, 1953. *Mathematics in Western Culture*, Oxford University Press.
- Kline, Morris, 1972. *Mathematical Thought From Ancient to Modern Times*, Oxford University Press.

- Knorr, Wilbur, 1975. *The Evolution of the Euclidean Elements*, Reidel, Boston.
- Knorr, Wilbur, 1976. "Problems in the interpretation of Greek number theory: Euclid and the 'fundamental theorem of arithmetic,'" *Studies in the Historical and Philosophical Sciences*, **7**, 353–368.
- Knorr, Wilbur, 1982. "Techniques of fractions in ancient Egypt," *Historia Mathematica*, **9**, No. 2, 133–171.
- Koblitz, Ann Hibner, 1983. *A Convergence of Lives. Sophia Kovalevskaja: Scientist, Writer, Revolutionary*. Birkhäuser, Boston.
- Koblitz, Ann Hibner, 1984. "Sofia Kovalevskaja and the mathematical community," *The Mathematical Intelligencer*, **6**, No. 1, 20–29.
- Koehler, Otto, 1937. *Bulletin of Animal Behavior*, No. 9. English translation in James R. Newman, ed., *The World of Mathematics*, vol. 1, Simon and Schuster, New York, 1956, pp. 491–492.
- Kowalewski, Gerhard, 1950. *Bestand und Wandel*, Oldenbourg, München.
- Kracht, Manfred; Kreyszig, Erwin, 1990. "E. W. von Tschirnhaus: his role in early calculus and his work and impact on algebra," *Historia Mathematica*, **17**, No. 1, 16–35.
- Kreyszig, Erwin, 1993. "On the calculus of variations and its major influences on the mathematics of the first half of our century," *Proceedings of the 19th Annual Meeting of the CSHPM/SCHPM*, Carleton University, Ottawa, Ontario, pp. 119–149.
- Kunoff, Sharon, 1990. "A curious counting/summation formula from the ancient Hindus," in: *Proceedings of the 16th Annual Meeting of the CSHPM/SCHPM*, University of Victoria, Victoria, British Columbia, pp. 101–107.
- Kunoff, Sharon, 1992. "Some inheritance problems in ancient Hebrew literature," in: *Proceedings of the 18th Annual Meeting of the CSHPM/SCHPM*, University of Prince Edward Island, Charlottetown, P.E.I., pp. 14–20.
- Lagrange, J.-L., 1771. "Réflexions sur la résolution algébrique des équations," *Nouveaux mémoires de l'Académie Royale des Sciences et Belles-lettres de Berlin; Œuvres* (1869), Vol. 3, pp. 205–421.
- Lagrange, J.-L., 1795. *Lectures on Elementary Mathematics*, translated by Thomas J. McCormack, Open Court, Chicago, 1898.
- Lam Lay-Yong, 1994. "Jiu Zhang Suanshu (*Nine Chapters on the Mathematical Art*): An Overview," *Archive for History of Exact Sciences*, **47**, No. 1, 1–51.
- Lam Lay-Yong; Ang Tian-Se, 1986. "Circle measurements in ancient China," *Historia Mathematica*, **13**, No. 4, 325–340.
- Lam Lay-Yong; Ang Tian-Se, 1987. "The earliest negative numbers: how they emerged from a solution of simultaneous linear equations," *Archive for History of Exact Sciences*, **37**, 222–267.
- Lam Lay-Yong; Ang Tian-Se, 1992. *Fleeting Footsteps. Tracing the Conception of Arithmetic and Algebra in Ancient China*, World Scientific, River Edge, NJ.
- Lam Lay-Yong; Shen Kangsheng, 1985. "The Chinese concept of Cavalieri's principle and its applications," *Historia Mathematica*, **12**, No. 3, 219–228.
- Lasserre, François, 1964. *The Birth of Mathematics in the Age of Plato*, Hutchinson, London.
- Lattin, Harriet Pratt, 1961. *The Letters of Gerbert*, Columbia University Press, New York.
- Laugwitz, Detlef, 1987. "Infinitely small quantities in Cauchy's textbooks," *Historia Mathematica*, **14**, No. 3, 258–274.
- Laugwitz, Detlef, 1999. *Bernhard Riemann, 1826–1866: Turning Points in the Conception of Mathematics*, Birkhäuser, Boston.
- Levey, Martin, 1966. *The Algebra of Abū Kāmil*, University of Wisconsin Press, Madison, WI.
- Levey, Martin; Petruck, Marvin, 1965. *Kūshyār ibn Labbān: Principles of Hindu Reckoning*, University of Wisconsin Press, Madison, WI.

- Libbrecht, Ulrich, 1973. *Chinese Mathematics in the Thirteenth Century*, MIT Press, Cambridge, MA.
- Liebmann, Heinrich, 1904. *N. J. Lobatschewskijs imaginäre Geometrie und Anwendung der imaginären Geometrie auf einige Integrale*, Teubner, Leipzig.
- Li Yan; Du Shiran, 1987. *Chinese Mathematics: A Concise History*, translated by John N. Crossley and Anthony W.-C. Lun, Clarendon Press, Oxford.
- Mack, John, 1990. *Emil Torday and the Art of the Congo. 1900–1909*, University of Washington Press, Seattle, WA.
- Maharajah, Bharati Krishna Tirthaji, 1965. *Vedic Mathematics*. Motilal Banarsidass Publishers, Delhi.
- Mallayya, V. Madhukar, 1997. “Arithmetic operation of division with special reference to Bhāskara II’s *Līlāvati* and its commentaries,” *Indian Journal of History of Science*, **32**, No. 4, 315–324.
- Mancosu, Paolo, 1989. “The metaphysics of the calculus: a foundational debate in the Paris Academy of Sciences, 1700–1706,” *Historia Mathematica*, **16**, No. 3, 224–248.
- Martzloff, Jean-Claude, 1982. “Li Shanlan (1811–1882) and Chinese traditional mathematics,” *The Mathematical Intelligencer*, **14**, No. 4, 32–37.
- Martzloff, Jean-Claude, 1990. “A survey of Japanese publications on the history of Japanese traditional mathematics (*Wasan*) from the last 30 years,” *Historia Mathematica*, **17**, No. 4, 366–373.
- Martzloff, Jean-Claude, 1993. “Eléments de réflexion sur les réactions chinoises à la géométrie euclidienne à la fin du XVIIIème siècle—Le *Jihe lunyue* {a} de Du Zhigeng {b} vue principalement à partir de la préface de l’auteur et deux notices bibliographiques rédigées par des lettrés illustres,” *Historia Mathematica*, **20**, 160–179.
- Martzloff, Jean-Claude, 1994. “Chinese mathematics,” in: I. Grattan-Guinness, ed., *Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences*, Vol. 1, Routledge, London, pp. 93–103.
- Matvievskaia, G. P., 1999. “On the Arabic commentaries to the tenth book of Euclid’s *Elements*,” *Istoriko-Matematicheskie Issledovaniya*, **4**, No. 39, 12–25 (Russian).
- Melville, Duncan, 2002. “Weighing stones in ancient Mesopotamia,” *Historia Mathematica*, **29**, No. 1, 1–12.
- Menninger, Karl, 1969. *Number Words and Number Symbols: A Cultural History of Numbers*, translated from the revised German edition by Paul Broneer. MIT Press, Cambridge, MA.
- Mikami Yoshio, 1913. *The Development of Mathematics in China and Japan*. Reprint: Chelsea, New York, 1961.
- Mikolás, M., 1975. “Some historical aspects of the development of mathematical analysis in Hungary,” *Historia Mathematica*, **2**, No. 2, 304–308.
- Milman, Dean; Guizot, M.; Smith, William, 1845. *The History of the Decline and Fall of the Roman Empire by Edward Gibbon*, John D. Morris, Philadelphia.
- Moore, Gregory H., 1982. *Zermelo’s Axiom of Choice*, Springer-Verlag, New York.
- Murata Tamotsu, 1994. “Indigenous Japanese mathematics, *wasan*,” in: I. Grattan-Guinness, ed., *Companion Encyclopedia of the History and Philosophy of Mathematical Science*, Vol. 1, Routledge, London, pp. 104–110.
- Narasimhan, Raghavan, 1990. *Bernhard Riemann: Gesammelte Mathematische Werke, Wissenschaftlicher Nachlaß und Nachträge*, Springer-Verlag, Berlin.
- Needham, J., 1959. *Science and Civilisation in China*, Vol. 3: *Mathematics and the Sciences of the Heavens and the Earth*, Cambridge University Press, London.
- Neugebauer, O., 1935. *Mathematische Keilschrifttexte*, Springer-Verlag, Berlin.
- Neugebauer, O., 1952. *The Exact Sciences in Antiquity*, Princeton University Press, Princeton, NJ.

- Neugebauer, O., 1975. *A History of Ancient Mathematical Astronomy* (three vols.), Springer-Verlag Berlin.
- Novosyolov, M. M., 2000. "On the history of the debate over intuitionistic logic," *Istoriko-Matematicheskie Issledovaniya*, **5**, No. 40, 272–280 (Russian).
- Ore, Oystein 1957. *Niels Henrik Abel, Mathematician Extraordinary*. Reprint: Chelsea, New York, 1974.
- Özdural, Alpay, 2000. "Mathematics and arts: connections between theory and practice in the Medieval Islamic world," *Historia Mathematica*, **27**, 171–200.
- Panteki, M., 1987. "William Wallace and the introduction of Continental calculus to Britain: a letter to George Peacock," *Historia Mathematica*, **14**, No. 2, 119–132.
- Parshall, Karen Hunger, 1985. "Joseph H. M. Wedderburn and the structure theory of algebras," *Archive for History of Exact Sciences*, **32**, Nos. 3/4, 223–349.
- Parshall, Karen Hunger, 1988. "The art of algebra from al-Khwarizmi to Viète: a study in the natural selection of ideas," *History of Science*, **26**, 129–164.
- Parshall, Karen Hunger, 2000. "Perspectives on American mathematics," *Bulletin of the American Mathematical Society*, **37**, No. 4, 381–405.
- Patterson, S. J., 1990. "Eisenstein and the quintic equation," *Historia Mathematica*, **17**, 132–140.
- Pavlov, Ivan, 1928. *Conditioned Reflexes*. Reprint: Dover, New York, 1960.
- Pavlov, Ivan, 1955. *Selected Works*, Foreign Languages Publishing House, Moscow.
- Perminov, V. Ya., 1997. "On the nature of deductive reasoning in the pre-Greek era of the development of mathematics," *Istoriko-Matematicheskie Issledovaniya* **2**, No. 37, 180–200 (Russian).
- Pesic, Peter, 2003. *Abel's Proof*, MIT Press, Cambridge, MA.
- Phili, Ch., 1997. "Sur le développement des mathématiques en Grèce durant la période 1850–1950. Les fondateurs," *Istoriko-Matematicheskie Issledovaniya*, 2nd ser., special issue on mathematical schools.
- Piaget, Jean, 1952. *The Child's Conception of Number*, Humanities Press, New York.
- Piaget, Jean; Inhelder, Bärbel, 1967. *The Child's Conception of Space*, Routledge & Kegan Paul, London.
- Picard, É., ed., 1897. *Œuvres Mathématiques d'Évariste Galois*, Gauthier-Villars, Paris.
- Pingree, David, 1968. "The fragments of the works of Ya'qūb ibn Tāriq," *Journal of Near Eastern Studies*, **26**, 97–125.
- Pingree, David, 1970. "The fragments of the works of al-Fazārī," *Journal of Near Eastern Studies*, **28**, 103–123.
- Pitcher, Everett, 1988. "The growth of the American Mathematical Society," *Notices of the American Mathematical Society*, **35**, No. 6, 781–782.
- Plofker, Kim, 2009. *Mathematics in India*. Princeton University Press.
- Poisson, S.-D., 1818. "Remarques sur les rapports qui existent entre la propagation des ondes à la surface de l'eau, et leur propagation dans une plaque élastique," *Bulletin des sciences, par la Société Philomatique de Paris*, 97–99.
- Price, D. J., 1964. "The Babylonian 'Pythagorean triangle'," *Centaurus*, **10**, 210–231.
- Pringsheim, A., 1910. "Über neue Gültigkeitsbedingungen für die Fouriersche Integralformel," *Mathematische Annalen*, **68**, 367–408.
- Rajagopal, P., 1993. "Infinite series in south Indian mathematics, 1400–1600," *Proceedings of the CSHPM/SCHPM 19th Annual Meeting*, Carleton University, Ottawa, Ontario, pp. 86–118.
- Rashed, Roshdi, 1989. "Ibn al-Haytham et les nombres parfaits," *Historia Mathematica*, **16**, No. 4, 343–352.

- Rashed, Roshdi, 1990. "A pioneer in anaclastics: Ibn Sahl on burning mirrors and lenses," *Isis*, **81**, No. 308, 464–491.
- Rashed, Roshdi, 1993. *Les mathématiques infinitésimales du IXe au XIe siècle*, Vol. II, Al-Furqan Islam Heritage Foundation, London.
- Rāshid, Rūshdī, 1994. *The Development of Arabic Mathematics: between Arithmetic and Algebra*, translated by Angela Armstrong, Kluwer Academic, Dordrecht and Boston.
- Reich, Karin, 1977. *Carl Friedrich Gauss: 1777/1977*, Inter Nationes, Bonn–Bad Godesberg.
- Richards, Joan, 1987. "Augustus de Morgan and the history of mathematics," *Isis*, **78**, No. 291, 7–30.
- Robins, Gay; Shute, Charles, 1987. *The Rhind Mathematical Papyrus: An Ancient Egyptian Text*, British Museum Publications, London.
- Robson, Eleanor, 1995. *Old Babylonian coefficient lists and the wider context of mathematics in ancient Mesopotamia 2100–1600 BC*. Dissertation, Oxford University.
- Robson, Eleanor, 1999. *Mesopotamian Mathematics, 2100–1600 BC: Technical Constants in Bureaucracy and Education*, Clarendon Press, Oxford and Oxford University Press, New York.
- Robson, Eleanor, 2001. "Neither Sherlock Holmes nor Babylon: A reassessment of Plimpton 322," *Historia Mathematica*, **28**, No. 3, 167–206.
- Robson, Eleanor, 2008. *Mathematics in Ancient Iraq*, Princeton University Press.
- Robson, Eleanor, 2009. "Mathematics education in an Old Babylonian scribal school," in: *Robson and Stedall, 2009*, pp. 199–227.
- Robson, Eleanor; Stedall, Jacqueline, eds., 2009. *The Oxford Handbook of the History of Mathematics*, Oxford University Press.
- Rosen, Frederic, 1831. *The Algebra of Mohammed ben Musa*, Oriental Translation Fund, London.
- Russell, Bertrand, 1945. *A History of Western Philosophy*, Simon and Schuster, New York.
- Sabra, A. I., 1969. "Simplicius's proof of Euclid's parallel postulate," *Journal of the Warburg and Courtauld Institute*, **32**, 1–24.
- Sabra, A. I., 1998. "One ibn al-Haytham or two? An exercise in reading the bio-bibliographical sources," *Zeitschrift für Geschichte der Arabisch-Islamischen Wissenschaft*, **12**, 1–50.
- Sarkor, Ramatosh, 1982. "The Bakhshali Manuscript," *Ganita-Bharati* (Bulletin of the Indian Society for the History of Mathematics), **4**, Nos. 1–2, 50–55.
- Scharlau, W., 1986. *Rudolf Lipschitz, Briefwechsel mit Cantor, Dedekind, Helmholtz, Kronecker, Weierstraß*, Vieweg, Deutsche Mathematiker-Vereinigung, Braunschweig–Wiesbaden.
- Scharlau, Winfried; Opolka, Hans, 1985. *From Fermat to Minkowski: Lectures on the Theory of Numbers and Its Historical Development*, Springer-Verlag, New York.
- Servos, John W., 1986. "Mathematics and the physical sciences in America, 1880–1930," *Isis*, **77**, 611–629.
- Sesiano, Jacques, 1982. *Books IV to VII of Diophantus' Arithmetica in the Arabic Translation Attributed to Qusta ibn Luqa*, Springer-Verlag, New York.
- Shen Kangshen, 1988. "Mutual-subtraction algorithm and its applications in ancient China," *Historia Mathematica*, **15**, 135–147.
- Siegmund-Schultze, Reinhard, 1988. *Ausgewählte Kapitel aus der Funktionenlehre*, Teubner, Leipzig.
- Siegmund-Schultze, Reinhard, 1997. "The emancipation of mathematical research publication in the United States from German dominance (1878–1945)," *Historia Mathematica*, **24**, 135–166.
- Sigler, L. E., transl., 1987. *Leonardo Pisano Fibonacci: The Book of Squares*, Academic Press, New York.

- Simonov, R. A., 1999. "Recent research on methods of rationalizing the computation of the Slavic Easter calculators (from manuscripts of the 14th through 17th centuries)," *Istoriko-Matematicheskie Issledovaniya*, **3**, No. 38, 11–31 (Russian).
- Singh, Parmanand, 1985. "The so-called Fibonacci numbers in ancient and medieval India," *Historia Mathematica*, **12**, No. 3, 229–244.
- Skinner, B. F., 1948. "'Superstition' in the pigeon," *Journal of Experimental Psychology*, **38**, No. 1 (February), 168–172.
- Smith, David Eugene, 1929. *A Source Book in Mathematics*, 2 vols. Reprint: Dover, New York, 1959.
- Smith, David Eugene; Ginsburg, Jekuthiel, 1934. *A History of Mathematics in America before 1900*, Mathematical Association of America/Open Court, Chicago (Carus Mathematical Monograph 5).
- Smith, David Eugene; Ginsburg, Jekuthiel, 1937. *Numbers and Numerals*, National Council of Teachers of Mathematics, Washington, DC.
- Smith, David Eugene; Latham, Marcia L., transl., 1954. *The Geometry of René Descartes*. Reprint: Dover, New York.
- Smith, David Eugene; Mikami, Yoshio, 1914. *A History of Japanese Mathematics*, Open Court, Chicago.
- Srinivasiengar, C. N., 1967. *The History of Ancient Indian Mathematics*, World Press Private, Calcutta.
- Stanley, Autumn, 1992. "The champion of women inventors," *American Heritage of Invention and Technology*, **8**, No. 1, 22–26.
- Stevin, Simon, 1585. *De Thiende*. German translation by Helmuth Gericke and Kurt Vogel. Akademische Verlag, Frankfurt am Main, 1965.
- Strauss, Walter, 1977. *Albrecht Dürer: The Painter's Manual*, Abaris Books, New York.
- Struik, D. J., 1933. "Outline of a history of differential geometry," *Isis*, **19**, 92–121, **20**, 161–192.
- Struik, D. J., ed., 1986. *A Source Book in Mathematics, 1200–1800*, Princeton University Press, Princeton, NJ.
- Stubhaug, Arild, 2000. *Niels Henrik Abel and his Times: Called Too Soon by Flames Afar*, Springer-Verlag, New York.
- Stubhaug, Arild, 2002. *The Mathematician Sophus Lie: It Was the Audacity of my Thinking*, translated from the Norwegian by Richard Daly, Springer-Verlag, Berlin.
- Swetz, Frank, 1977. *Was Pythagoras Chinese? An Examination of Right Triangle Theory in Ancient China*, The Pennsylvania State University Studies, No. 40. The Pennsylvania State University Press, University Park, PA, and National Council of Teachers of Mathematics, Reston, VA.
- Thomas, Ivor, 1939. *Selections Illustrating the History of Greek Mathematics*, Vol. 1, *Thales to Euclid*, Harvard University Press, Cambridge, MA.
- Thomas, Ivor, 1941. *Selections Illustrating the History of Greek Mathematics*, Vol. 2, *Aristarchus to Pappus*, Harvard University Press, Cambridge, MA.
- Thoren, Victor E., 1988. "Prosthaphæresis revisited." *Historia Mathematica*, **15**, 32–39.
- Todhunter, Isaac, 1861. *A History of the Calculus of Variations in the Nineteenth Century*. Reprint: Chelsea, New York, 1962.
- Todhunter, Isaac, 1865. *A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace*. Reprint: Chelsea, New York, 1949.
- Toomer, G. J., 1976. *Diocles on Burning Mirrors: The Arabic Translation of the Lost Greek Original*, Springer-Verlag, New York.
- Toomer, G. J., 1984a. *Ptolemy's Almagest*, Springer-Verlag, New York.

- Toomer, G. J., 1984b. "Lost Greek mathematical works in Arabic translation," *The Mathematical Intelligencer*, **4**, No. 2, 32–38.
- Toomer, G. J., 1990. *Conics, Books V to VII. The Arabic Translation of the Lost Greek Original in the Version of the Banu Musa*. Springer, New York.
- Tropfke, 1902. *Geschichte der Elementarmathematik*, 4. Auflage, completely revised by Kurt Vogel, Karin Reich, and Helmuth Gericke. Band 1: *Arithmetik und Algebra*. Walter de Gruyter, Berlin and New York, 1980.
- Tsaban, Boaz; Garber, David, 1998. "On the rabbinical approximation of π ," *Historia Mathematica*, **25**, No. 1, 75–84.
- Unguru, Sabetai, 1975/76. "On the need to rewrite the history of Greek mathematics," *Archive for History of Exact Sciences*, **15**, No. 1, 67–114.
- Urton, Gary, 1997. *The Social Life of Numbers: A Quechua Ontology of Numbers and Philosophy of Arithmetic*, University of Texas Press, Austin, TX.
- Varadarajan, V. S., 1983. "Mathematics in and out of Indian universities," *The Mathematical Intelligencer*, **5**, No. 1, 38–42.
- Vidyabhusana, Satis Chandra, 1971. *A History of Indian Logic (Ancient, Mediaeval, and Modern Schools)*, Motilal Banarsidass, Delhi.
- Volkov, Alexei, 1997. "Zhao Youqin and his calculation of π ," *Historia Mathematica*, **24**, No. 3, 301–331.
- van der Waerden, B. L., 1963. *Science Awakening*, Wiley, New York.
- van der Waerden, B. L., 1975. "On the sources of my book *Moderne Algebra*," *Historia Mathematica*, **2**, 31–40.
- van der Waerden, B. L., 1983. *Geometry and Algebra in Ancient Civilizations*, Springer-Verlag, New York.
- van der Waerden, B. L. 1985. *A History of Algebra from al-Khwarizmi to Emmy Noether*, Springer-Verlag, Berlin.
- Walford, E., transl., 1853. *The Ecclesiastical History of Socrates, Surnamed Scholasticus. A History of the Church in Seven Books*, H. Bohn, London.
- Wantzel, Laurent, 1837. "Recherches sur les moyens de reconnaître si un problème de Géométrie peut se résoudre avec la règle et le compas," *Journal de mathématiques pures et appliquées*, **2**, 366–372.
- Wantzel, Laurent, 1843. "Classification des nombres incommensurables d'origine algébrique," *Nouvelles annales de mathématiques*, **2**, 117–127.
- Wantzel, Laurent, 1845. "Démonstration de l'impossibilité de résoudre toutes les équations algébriques avec des radicaux," *Bulletin des sciences, par la Société Philomathique de Paris*, 5–7.
- Waring, E. 1762. *Miscellanea analytica*, Oxford.
- Weil, André, 1984. *Number Theory: An Approach Through History from Hammurapi to Legendre*, Birkhäuser, Boston.
- Whiteside, T. L., 1967. *The Mathematical Papers of Isaac Newton*, Johnson Reprint Corporation, London.
- Witmer, T. Richard, transl., 1968. *The Great Art, or The Rules of Algebra, by Girolamo Cardano*, MIT Press, Cambridge, MA.
- Woepcke, Franz, 1852. "Notice sur une théorie ajoutée par Thâbit ben Korrah à l'arithmétique spéculative des Grècs," *Journal asiatique*, **4**, No. 20, 420–429.
- Woodhouse, Robert, 1810. *A History of the Calculus of Variations in the Eighteenth Century*. Reprint: Chelsea, New York, 1964.

- Yoshida Kôzaku, 1980. "Mathematical works of Takakazu Seki," *The Mathematical Intelligencer*, **3**, No. 3: Reprint of material written for the centennial of the Japan Academy.
- Zaitsev, E. A., 1999. "The meaning of early medieval geometry: from Euclid and surveyors' manuals to Christian philosophy," *Isis*, **90**, 522–553.
- Zaitsev, E. A., 2000. "The Latin versions of Euclid's *Elements* and the hermeneutics of the twelfth century," *Istoriko-Matematicheskie Issledovaniya*, **5**, No. 40, 222–232 (Russian).
- Zeuthen, H. G., 1903. *Geschichte der Mathematik im 16. und 17. Jahrhundert*, Teubner, Stuttgart. Johnson Reprint Corporation, New York, 1966.
- Zharov, V. K., 2001. "On the 'Introduction' to the treatise *Suan Shu Chimeng* of Zhu Shijie," *Istoriko-Matematicheskie Issledovaniya*, 2nd Ser., **6**, No. 41, 347–353 (Russian).
- Zhmud, Leonid, 1989. "Pythagoras as a mathematician," *Historia Mathematica*, **16**, No. 3, 249–268.
- Zverkina, G. A., 2000. "The Euclidean algorithm as a computational procedure in ancient mathematics," *Istoriko-Matematicheskie Issledovaniya*, **5**, No. 40, 232–243 (Russian).

NAME INDEX

- A-user-re, 58
Abel, Niels Henrik, 4, 5, 382, 400, 407, 442–443, 446, 447, 502, 526
Absolon, Karel, 16
Abu Kamil, 288, 292, 297, 300, 324
Abu'l Wafa, 288, 295
Abu'l-Wafa, 292
Achilles, 107
Adalbold of Liège, 316
Adelard of Bath, 318
Adler, I., 325
Ağargün, Ahmet, 294
Agnesi, Maria Gaetana, 382, 399, 405
Ahmose, 58, 76
Akademos, 135
Akbar the Lion, 204, 210
Albategnius, 288
Alberti, Leon Battista, 321, 352
Aleksandrov, Pavel Sergeevich, 535
d'Alambert, Jean le Rond, 386, 390, 397, 437–440, 512, 513
Alexander of Macedon, 28, 57, 79, 116, 134, 178
Alexander Polyhistor, 84
Alhazen (ibn al-Haytham), 288, 305
Allman, George Johnston, 47, 123, 130
Althoff, Friedrich, 409
Amenemhet III, 58
Amenhotep III, 68
Amir-Moez, Ali R., 297, 298, 307, 308
Ampère, André-Marie, 511, 527
Anaxagoras, 83, 115, 117
Ang Tian-Se, 242, 243, 245, 250
Anne, British queen, 376
Antiphon, 116, 131
Apepi I, 58
Apollodorus, 84
Apollonius, xxv, 8, 29, 77, 79, 83, 88, 89, 105, 119, 140, 160–169, 172, 176, 177, 192, 199, 262, 284, 298, 305, 359, 492
Archimedes, xxv, 8, 29, 77, 79–81, 83, 87, 118, 131, 140, 148–160, 169, 171, 172, 176, 177, 194, 211, 235, 244, 262, 263, 266, 279, 284, 303, 305, 319, 360, 365, 367, 371
Archytas, 83, 98, 118, 128, 130, 210
Argand, Jean, 499
Aristaeus, 160, 161
Aristarchus, 189
Aristotle, 77, 80, 81, 83, 84, 107, 109, 111, 113, 116, 128–139, 171, 293, 307, 417, 425, 476
Arjuna, 233
Artin, Emil, 413
Aryabhata I, 207, 211, 219–227, 230, 235, 250, 284, 288, 304
Ascher, Marcia, 17, 19
Averroes, 291
Avicenna, 291
Avogadro, Amadeo, 430
Ayoub, Raymond, 441, 442
Azulai, Abraham, 294

Baer, Nicolai Reymers, 334
Bagheri, Muhammad, 287
Baigozhina, G. O., 293
Baire, René, 529, 538, 539
Ball, W. W. Rouse, 203, 409
Baltzer, R., 452
Banach, Stefan, 540
Barabe, D., 325
Barrow, Isaac, 370, 373, 376, 382
Bartels, Johann, 488
Barzin, M., 553
Bashmakova, I. G., 134
al-Battani, 288, 318

- Bekker, August Immanuel, 56, 109, 293
 Belcastro, Sarah-Marie, 18
 Beltrami, Eugenio, 478, 490
 Belyi, Andrei, 529
 Beman, W. W., 126
 Bendixson, Ivar, 535
 Bentham, Jeremy, 543
 Berggren, J. L., 286, 290, 305
 Berkeley, George, 379, 382, 385
 Bernal, Martin, 80
 Bernays, Paul, 492, 554
 Berndt, Bruce, 211
 Bernoulli, Daniel, 376, 390, 394, 424,
 511–513
 Bernoulli, James, 379, 393, 421–424, 428,
 431, 501
 Bernoulli, John, 379, 389, 392, 437,
 495, 536
 Bernoulli, Nicholas I, 437
 Bernoulli, Nicholas II, 424
 Bessel, Friedrich Wilhelm, 486, 488
 Betti, Enrico, 444, 478
 Bézout, Etienne, 438, 439, 463
 Bhaskara I, 216
 Bhaskara II, 209–211, 233–238, 266,
 275, 324
 Bhau Daji, 207
 Bianchi, Luigi, 479
 Bieberbach, Ludwig, 414
 Bienaimé, Irénée-Jules, 428
 al-Biruni, 172, 206, 289, 290, 305
 Bochenski, I. M., 543
 Boethius, 96, 313, 316
 Bólyai, Farkas, 485, 487
 Bólyai, János, 456, 487, 489
 Bolzano, Bernard, 499, 517, 525, 527
 Bombelli, Rafael, 322, 340, 342, 496, 498
 Boncompagni, Baldassare, 319, 327
 Bonnet, Ossian, 473
 Boole, George, 542, 544, 546–548, 550
 Borel, Emile, 526, 535, 538
 Bosse, Abraham, 354
 Bottazzini, Umberto, 438, 514, 526, 527
 Bouvelles, Charles, 364
 Bouvet, Joachim, 244
 Boyer, Carl, 331, 363
 Brahe, Tycho, 334
 Brahmagupta, 19, 207–209, 224, 227–238, 254,
 284, 324
 Brandt, Heinrich, 414
 Bravais, Auguste, 325
 Bravais, Louis, 325
 Brea, Bernabò, 149
 Brentjes, Sonja, 293
 Bretschneider, Carl Anton, 47
 Brett, William F., 290
 Brianchon, Charles, 449
 Briggs, Henry, 344
 Brontinus, 197
 Brouwer, L. E. J., 551, 556
 Bruins, Evert Marie, 42
 Buck, R. C., 54
 Buddha, Gautama, 204
 Bugaev, Nikolai Vasilevich, 529
 Burali-Forti, Cesare, 539, 549
 Bürgi, Jobst, 333
 Burkert, Walter, 83
 Butzmann, Hans, 175
 Bychkov, S. N., 82
 Callahan, Jeremiah J., 494
 Cantor, Georg, 512, 528, 532–541, 543,
 546, 549
 Cantor, Moritz, 71, 72, 294
 Cardano, Girolamo, 94, 321, 338, 342,
 418–419, 428, 429, 435
 Carleson, Lennart, 527
 Carnot, Lazare, 353
 Castor, 135
 Catherine II, 386
 Cauchy, Augustin-Louis, 387, 391, 398–400,
 441–442, 452, 458, 504–506, 514, 516,
 521
 Cavalieri, Bonaventura, 195, 365, 368, 369, 382
 Cayley, Arthur, 408, 455–456, 463, 546
 Cebes, 86
 Chace, A. B., 58
 Chaikovskii, Yu. V., 418
 Chandrasekharan, Komaravolu, 210
 Charlemagne, 284, 315, 318
 Charles II, 370, 373
 Charles Martel, 283
 Chasles, Michel, 351
 Chebyshev, Pafnutii L'vovich, 428, 430
 Cheng Dawei, 243, 269
 Chiang Kai-Shek, 240
 Christianidis, Jean, 101
 Chuquet, Nicolas, 320, 335, 343
 Ciancio, Salvatore, 149
 Cicero, 81, 148
 Clagett, Marshall, 149, 331
 Clark, Walter Eugene, 208, 219, 220, 222, 224
 Clavius, Christopher, 244
 Clebsch, Alfred, 506

- Clement of Alexandria, 68
 Cleopatra, 29
 Clifford, William Kingdon, 351
 Closs, Michael, 197
 Codazzi, Delfino, 478
 Cohen, Paul, 536
 Colebrooke, Henry Thomas, 208–210, 227,
 228, 233, 285, 294
 Collins, James, 376, 381
 Columbus, Christopher, 323
 Confucius, 239
 Conon, 149, 150
 Constantine of Fleury, 316
 Coolidge, Julian Lowell, 351, 449, 464, 468,
 482, 489, 492
 Copernicus, Nicolaus, 183, 309
 Cossali, Pietro, 295
 Cotes, Roger, 379, 424
 Courant, Richard, 413
 Craik, Alex D., 381
 Cramér, Gabriel, 453
 Croesus, 82
 Crossley, John N., 294
 Cullen, Christopher, 241, 249
 Cuomo, Serafina, 196
 Cyril, 197, 198
 Cyrus the Great, 28, 29

 D'ooge, Martin Luther, 96
 al-Daffa, Abu Ali, 292, 295
 Damo, 197
 Dante, 315, 317, 323
 Darboux, Gaston, 391, 514
 Darius, 29, 178
 David, H. A., 431
 Davis, Philip, 360, 362
 De Moivre, Abraham, 423–424, 428, 429
 De Mora-Charles, S., 421
 De Morgan, Augustus, 244, 542–545, 550
 De Young, Gregg, 210, 290
 Deakin, Michael, 198
 Dedekind, Richard, 461, 523–525, 534, 537
 Degen, Ferdinand, 442
 Delamain, Richard, 345
 Democritus, 68
 Demosthenes, 81
 Denjoy, Arnaud, 526
 Desargues, Girard, 192, 352–355, 451
 Descartes, René, 166, 191, 288, 306, 354,
 359–362, 364, 372, 376, 382, 393, 448,
 453, 456, 511, 521, 537
 Detlefsen, Michael, 19, 555

 Devlin, Keith, 319, 324
 Dick, Auguste, 412, 413
 Dickson, L. E., 94, 229, 247, 293, 443
 Diels, Hermann, 86
 Dijksterhuis, E. J., 172
 Dilke, O. A. W., 174, 179
 Dinostratus, 117, 123
 Diocles, 169
 Diocletian, 190
 Diogenes Laertius, 81, 82, 84, 123, 128, 130,
 134, 197
 Dion, 128
 Dionysus I, 128
 Diophantus, 20, 77, 81, 89, 91, 102, 198, 213,
 219, 229, 230, 254, 292, 295, 322, 327, 341
 Dirichlet, Peter, 515, 518, 525
 Djoser, 56
 Dorofeeva, A. V., 395
 Dositheus, 149, 150
 Du Bois-Reymond, Paul, 527, 529
 Du Shiran, 243, 244, 249, 252, 255–257, 284
 Du Zhigeng, 244
 Dudley, Underwood, 294
 Duillier, Nicolas Fatio de, 381
 Dummit, David, 445
 Dupin, Pierre, 468
 Dürer, Albrecht, 349, 466
 Dzielska, Maria, 198

 Edward VI, 340
 Edwards, A. W. F., 431
 Egorov, Dmitrii Fyodorovich, 529
 Ehara Masanori, 277
 Einstein, Albert, 412, 476
 Emperor Yu, 250
 Eratosthenes, 83, 93, 117, 149, 154, 177,
 179, 317
 Erdmann, G., 397
 Erdős, Pál, 294
 Errera, A., 553
 Esau, 294
 Euclid, xxv, 7, 13, 21, 29, 57, 77, 81, 83, 86, 87,
 89, 91, 92, 94, 97, 98, 103, 112, 113, 116,
 117, 130, 136, 140–147, 149, 160, 169,
 171, 172, 176, 177, 185, 192, 193, 196,
 210, 214, 238, 243, 262, 284, 288, 290,
 293, 298, 302–309, 313, 318, 327, 348,
 351, 356, 362, 373, 481, 483, 491, 499,
 521, 524
 Eudemus, 81, 83, 105, 116, 123, 161
 Eudoxus, 79, 83, 87, 116, 128–130, 134, 141,
 144, 160, 176, 194, 195, 384, 397

- Euler, Leonhard, 17, 95, 97, 100, 229, 376, 382, 386, 390, 393–394, 423, 437–439, 453, 468, 470, 473, 476, 479, 495, 498–500, 502, 512, 513, 522, 530
- Eutocius, 81, 118, 128, 148, 160, 161, 171
- di Fagnano, Giulio de' Toschi, 501
- al-Farisi, Kamal al-Din, 294
- Fawcett, Philippa, 409
- Fechner, Gustave Theodor, 425
- Feigenbaum, L., 388
- Feingold, Mordechai, 382
- Feldman, Emile B., 290
- de Fermat, Pierre, 100, 166, 191, 229, 359, 361, 363, 364, 367, 369, 372, 376, 382, 419–421, 521
- Ferrari, Ludovico, 322, 339
- Ferreirós, José, 534
- del Ferro, Scipione, 321
- Fibonacci, 288, 297, 319
- Field, J. V., 137, 352, 355, 451
- Finzi, Mordecai, 297
- Fior, Antonio Maria, 321
- Fischer, Ernst, 412
- Fitzgerald, Augustine, 197
- Flegg, G., 320
- Fletcher, Colin, 294
- Fontana, Niccolò, 321
- de Fontenelle, Bernard Lebouyer, 381
- Fourier, Joseph, 377, 391, 473, 514, 516
- Fowler, David, 129
- Fraenkel, Adolf, 554
- della Francesca, Piero, 349
- Franchella, Miriam, 551
- Franci, Rafaella, 335
- Fraser, Craig, 395, 398
- Frederick I, 319
- Frederick II, 319, 324, 326
- Frege, Gottlob, 548
- Frenet, Jean, 477
- Friberg, Jöran, 52, 53
- Fried, Michael N., 160, 161, 165
- von Frisch, Karl, 16
- von Fritz, Kurt, 111
- Fu Daiwie, 264
- Fubini, Guido, 541
- Fukagawa Hidetoshi, 278
- Fuson, Karen, 15
- Galileo, 178, 365, 465
- Galois, Evariste, 4, 5, 443–447
- Gandz, Solomon, 294
- Garber, David, 70
- Gardner, Milo, 58
- al-Gauhari, 302
- Gauss, Carl Friedrich, 247, 426–428, 439–440, 456, 469–473, 475, 477, 485–486, 488, 489, 493, 499, 502, 505, 532, 544
- Gellius, Aulus, 84
- Gelon, 149
- Geminus, 161, 171
- Genghis Khan, 204, 240, 290
- George I, 376, 469
- George IV, 469
- Gerbert, 315–318, 323
- Gerbillon, Jean-François, 244
- Gerdes, Paulus, 71
- Gergonne, Joseph, 452
- Gerhardt, C. I., 388, 523
- Gericke, Helmut, 329
- Gerling, Christian Ludwig, 485
- Germain, Sophie, 473
- Gherard of Cremona, 318
- Gibbon, Edward, 198
- Gillings, Richard, 63, 67, 71, 74
- Ginsburg, Jekuthiel, 31
- Girard, Albert, 434, 436, 446
- Glivenko, Valerii Ivanovich, 553
- Gödel, Kurt, 536, 542, 549, 554
- Goldbach, Christian, 9
- Goldstine, Herman, 392
- Golenishchev, Vladimir Semënovich, 57
- Goodwin, Edwin J., 70
- Gordan, Paul, 412
- Gould, Carol Grant, 16
- Gould, James L., 16
- Goursat, Edouard, 508
- Gow, James, 90
- Grabiner, Judith, 360, 399
- Grassmann, Hermann, 473–474
- Grattan-Guinness, Ivor, 3, 408, 410, 452, 514, 535, 537, 545, 548, 550
- Gray, J. J., 105, 113, 136, 137, 302, 306–308, 352, 355, 451, 490, 492
- Green, George, 505
- Greenleaf, Benjamin, 253
- Gregory VII, 318
- Gregory, James, 370, 374, 523
- Grosholz, Emily, 387
- Guldin, Habakuk Paul, 195
- Guo Shuchun, 242
- Hadamard, Jacques, 538, 539
- Hairetdinova, N. G., 308

- Halayudha, 217
Halley, Edmund, 161
Hamilton, William Rowan, 5, 442, 473, 479
Hammurabi, 28
Hankel, Hermann, 529
Hardy, G. H., 211
Harish-Chandra, 210
Harnack, Axel, 526
Harriot, Thomas, 393
al-Hassar, Abu Bakr, 216
ibn al-Haytham, 288, 293, 302, 305, 309, 310, 393, 479, 484
He Chengtian, 264
Heath, T. L., 87, 99, 160, 172, 190
Hector, 107
Heiberg, Johann Ludwig, 153
Helen, 135
Helicon, 130
Henri IV, 322, 341, 423
Henry, Alan S., 294
Hensel, Kurt, 530
Heracleides, 148, 160
Heraclitus, 85
Herbart, Johann Friedrich, 451, 475
Hermite, Charles, 445, 523
Hermodorus, 89
Herodotus, 27, 68, 82
Heron, 78, 88, 98, 154, 172–174, 176, 177
Hersh, Reuben, 360, 362, 543
Heyde, C. C., 429
Heytesbury, William, 331
Hideyoshi, 268
Hieron II, 148, 149
Hieronymus of Rhodes, 82
Hilbert, David, 412, 491, 536, 537, 550
Hille, Einar, 433, 499
Hipparchus, 79, 130, 177, 183, 197
Hippasus, 84
Hippias, 117, 123, 169
Hippocrates, 116–118, 132, 138, 144
Hobson, E. W., 526
Hogendijk, Jan, 263, 293, 299
Holt, P.M., 318
Homann, Frederick A., 317
Homer, 107
Horner, William, 258
Horus, 62, 72
de l'Hospital, Marquis, 378–380, 382
Høyrup, Jens, 38, 46
Hughes, Barnabas, 329
Hulegu, 290
Huygens, Christiaan, 393, 420–421, 424, 456, 464–466, 479
Hypatia, 78, 79, 81, 89, 99, 190, 196–199, 405
Iamblichus, 81, 84, 293
Il'ina, E. A., 134
Indorato, Luigi, 364
Inhelder, Bärbel, 17
Innocent III, 319
Isis, 72
Isodoros, 190, 198
Isomura Kittoku, 275, 280
Ivins, W. M., 352
Jacob, 294
Jacobi, Carl Gustav, 97, 387, 395, 397, 407, 444, 478, 502
al-Jayyani, 308
Jean, R. V., 325
Jentsch, Werner, 414
Jerome, 81
Jevons, William Stanley, 548
Jia Xian, 257
John of Palermo, 326, 327
Jones, Alexander, 79, 87, 140
Jordanus Nemorarius, 319, 329
Julius Caesar, 29, 81
Justinian, 291
Jyesthadeva, 227, 358, 371, 374
Kang Xi, 244
Kant, Immanuel, 7, 385, 541
al-Karaji, 327
al-Kashi, 240, 286, 292
Kasir, Daoud, 289, 298, 308
Katsma, Robert W., 466
Kazdan, Jerry L., 244
Kepler, Johannes, 183, 315, 372, 418, 451
Khafre, 70
al-Khayyam, Umar, 289
Khayyam, Omar, 297–299, 307, 479
Khinchin, Aleksandr Yakovlevich, 553
Khufu, 82
al-Khwarizmi, 286, 287, 289, 290, 294–297, 300, 318, 324, 329, 335
King, R. Bruce, 445
Kingsley, Charles, 198
Kiro, S. N., 488
Kitagawa Mōko, 277
Klein, Felix, xxiii, 126, 353, 408, 409, 412, 443, 445, 450, 451, 454, 456, 472, 475, 476, 478, 485, 486, 491, 504, 537

- Kline, Morris, 291
 Kneser, Adolf, 395
 Knorr, Wilbur, 97, 110, 111, 113, 123
 Kobayashi, Shōshichi, 270
 Koehler, O., 14
 Kovalevskaya, Sof'ya Vasil'evna, 391, 406–408, 412, 414
 Kovalevskaya, Sof'ya Vladimirovna, 407
 Kovalevskii, Vladimir Onufrevich, 406
 Kowalewski, Gerhard, 412
 Kracht, Manfred, 436
 Kreyszig, Erwin, 392, 436
 Kronecker, Leopold, 532, 534, 536, 537, 543
 Kryukovskaya, Anna Vasil'evna, 406
 Kublai Khan, 240
 al-Kuhi, 305
 Kummer, Ernst Eduard, 534, 536

 ben Laban, Kushar, 292
 Lagrange, Joseph-Louis, 229, 386, 389, 394, 398, 400, 437–440, 446, 469, 500, 510, 528, 545
 Lam Lay-Yong, 242, 245, 252, 264
 Lambert, Johann, 303, 484–485, 487
 Lambton, Ann K. S., 318
 Lamé, Gabriel, 473
 Lander, L. J., 102
 Lao-Tzu, 239
 Laplace, Pierre-Simon, 425, 428, 513, 516
 Lasserre, François, 86, 129
 Latham, Marcia L., 361
 Lattin, Harriet, 316
 Laugwitz, Detlef, 399, 475, 477
 Laurent, Pierre, 505
 Lebesgue, Henri, 515, 526, 528, 538, 539
 Lebesgue, Victor-Amédée, 95
 Legendre, Adrien-Marie, 386, 394, 426, 485, 502, 503
 Lehmus, D. C. L., 199
 Leibniz, Gottfried, 9, 10, 216, 273, 312, 345, 355, 358, 373–385, 387, 388, 393, 397, 420–421, 429, 435–437, 453, 456, 467–469, 495, 500, 522, 542, 543, 549
 Leon, 98
 Leonardo of Pisa, 288, 297, 316, 319, 324–328, 335, 336, 341, 349
 Leonidas, 115
 Levey, Martin, 324
 Levi-Civita, Tullio, 478
 Lewis, Bernard, 318
 Li Ang, 256
 Li Shanlan, 244
 Li Yan, 243, 244, 249, 252, 255, 257, 284
 Li Ye, 271
 Li Zhi, 271
 Libbrecht, Ulrich, 261, 271
 Liebmann, Heinrich, 488, 489
 Lincoln, Abraham, 21, 31
 Lindemann, Ferdinand, 445, 523
 Liouville, Joseph, 472, 473, 515, 523
 Lipschitz, Rudolf, 524
 Listing, Johann Benedict, 456
 Liu Hui, 202, 242, 243, 250, 252, 262–266
 Lobachevskii, Nikolai Ivanovich, 456, 486–489
 Loomis, Elias, 244
 Loria, Gino, 408, 410, 415
 Louis XIV, 244, 376, 423
 Lull, Ramon, 19, 360
 Luzin, Nikolai Nikolaevich, 527, 528, 534, 535, 540
 Lysis, 197

 Maclaurin, Colin, 382, 388, 453, 463
 Maddison, Isabel, 409
 Madhava, 358, 371
 Maharaja, Bharati, 205
 Mahavira, 204, 206
 Maheswara, 209
 Mainardi, Gaspare, 477
 al-Majriti, 294
 Maltby, Margaret Eliza, 409
 al-Mamun, 286, 287, 295
 Mancosu, Paolo, 381, 397
 Mann, Thomas, 516
 al-Mansur, 284, 285, 292
 Marcellus, 148
 Marco Polo, 240
 Marinus of Tyre, 177, 179
 Markov, Andrei Andreevich, 429
 Martzloff, Jean-Claude, 242, 244, 267
 Master Hugh, 317
 Matvievskaya, G. P., 302
 de' Mazzinghi, Antonio, 335
 Melville, Duncan, 29
 Menaechmus, 83, 116, 118, 127, 128
 Menciuss, 239
 Mencke, Otto, 377
 Menelaus, 123, 184, 351, 353
 Menes, 56
 Menna, 68
 Menninger, Karl, 31, 207
 de Méré, Chevalier, 419
 Mersenne, Marin, 95, 354, 363

- metaphysics
 Pythagorean, 111
 Mikami Yoshio, 239, 244, 256, 267, 271–273, 275, 279
 Milnor, John, 479
 Minding, Ferdinand, 473, 486
 Minos, 118
 Mittag-Leffler, Gösta, 407
 Möbius, August Ferdinand, 449, 452–453, 458–459
 Monbu, 268
 Monge, Gaspard, 450–451, 469, 473
 Monk, J. Donald, 554
 Moore, Gregory, 538, 549
 Mōri Kambei, 268
 Mōri Shigeyoshi, 268
 Morley, Frank, 199
 Muir, Thomas, 270
 Müller, Johann, 309, 320
 Murata Tamotsu, 268, 280
- Nachshon, Rau, 294
 Napier, John, 322, 343–345
 Napoleon, 451
 Narasimhan, Raghavan, 475, 477
 Narmer, 56
 Naucrates, 88
 al-Nayrizi, Abu, 136
 Nebuchadnezzar, 28
 Nehru, Jawaharlal, 210
 Nesselmann, G. H. F., 99
 Neugebauer, Otto, 29, 39, 41, 43, 46, 48, 49, 52, 73, 86, 98
 Newson, Henry, 410
 Newton, Isaac, 8, 82, 157, 309, 312, 358, 362, 373–385, 388, 393, 424, 436–437, 448, 451, 453, 466–467, 476, 479, 521, 523
 Nicomachus, 77, 91, 92, 238, 293, 313
 Nicomedes, 123, 126, 169
 Nieuwentijt, Bernard, 378
 Nilakanta, 358
 Nipsus, M. Iunius, 174
 Noether, Emmy, 411–414
 Noether, Fritz, 412, 414
 Noether, Max, 412
 Novosyolov, M. M., 553
 Ny-maat-re, 58
- O'Connor, J. J., 196
 Oldenburg, Henry, 371, 376, 435
 Omar Khayyam, 289, 290, 297, 321
 d'Oresme, Nicole, 288, 319, 330, 335, 337, 361, 383
 Orestes, 197
 Osiris, 72
 Ostrogradskii, Mikhail Vasilevich, 505
 Otto III, 316
 Oughtred, William, 345
 Özdural, Alpay, 285
- Pacioli, Luca, 320, 335, 349
 Padmanabha, 209
 Pamphila, 82
 Panini, 205
 Panteki, M., 381
 Pappus, 8, 78, 81, 88, 89, 105, 123, 126, 140, 149, 160, 161, 166, 190–199, 297, 353, 358, 361, 456
 Parkin, T. R., 102
 Parmenides, 85
 Parshall, Karen, 335
 Pascal, Blaise, 191, 218, 345, 355, 368–370, 376, 382, 419–421, 429, 434, 462, 463
 Pavlov, Ivan Petrovich, 16, 20, 427
 Peano, Giuseppe, xxiii, 549
 Pearson, Karl, 424
 Peaucellier, Charles-Nicolas, 124
 Peet, T. E., 73
 Peirce, Charles Sanders, 543, 556
 Pell, John, 229
 Perelman, Grigorii, 461
 Pericles, 115, 117
 Perminov, V. Ya., 22
 Perron, Oskar, 526
 Pestalozzi, Johann Heinrich, 451
 Peter I, 376
 Peterson, Karl Mikhailovich, 477
 Philip of Macedon, 28, 116
 Philolaus, 86, 115, 128, 129
 Piaget, Jean, 16, 17
 Picard, Emile, 444
 Pincherle, Salvatore, 527
 Pingala, 205, 217
 Pingree, David, 286
 Pitiscus, Bartholomeus, 332–334, 337
 Planudes, Maximus, 101
 Plato, 47, 71, 77, 80, 81, 83, 85, 90, 96, 97, 110, 111, 115, 117, 128–139, 171, 188, 210, 315, 359, 417
 Plato of Tivoli, 318
 Playfair, John, 171
 Pliny the Elder, 178

- Plofker, Kim, 205, 213, 216, 219, 222, 235,
 285, 286, 305
 Plücker, Julius, 454–455, 462
 Plutarch, 71, 81, 82, 113, 117, 128, 130,
 148, 149
 Poincaré, Henri, 459–461, 478, 491, 537
 Poisson, Siméon-Denis, 427, 428, 473, 516, 517
 Pollux, 135
 Polybius, 178
 Poncelet, Jean-Victor, 451
 Price, D. J., 51
 Pringsheim, Alfred, 516
 Proclus, 81, 83, 85, 86, 96, 103, 105, 113, 116,
 119, 123, 126, 130, 136, 149, 161, 171, 482
 Psellus, Michael, 98
 Pseudo-Boethius, 314
 Ptolemy (Egyptian ruler), 118
 Ptolemy Euergetes, 160
 Ptolemy Soter, 29, 87, 116
 Ptolemy, Claudius, 8, 33, 77, 79, 81–83, 89,
 136, 161, 172, 174, 177–190, 220, 225,
 238, 286, 290, 304, 306, 308, 315, 317,
 320, 482, 484
 Puiseux, Victor, 458, 506
 Pythagoras, 46, 80, 82–85, 97, 104, 197,
 293, 315
 Pytheas, 178

 Qin Jiushao, 258, 260, 271
 ibn-Qurra, Thabit, 287, 293, 294, 302, 309, 318,
 482, 484, 493

 Rajagopal, P., 358
 Ramanujan, Srinivasa, 210, 244
 Ramses II, 68
 al-Raschid, Harun, 284
 Rashed, Roshdi, 305, 306
 R^{ashid}, Rushdⁱ, 288
 Rawlinson, Sir Henry, 29
 Recorde, Robert, 328, 340
 Regiomontanus, 309, 320, 331, 337
 Reich, Karen, 426, 469
 Reinhardt, Curt, 459
 Reisner, George Andrew, 57
 Rhind, Alexander Henry, 57
 Ricci, Matteo, 240, 243
 Ricci-Curbastro, Gregorio, 478
 Richards, Joan, 543
 Richer, 316
 Riemann, Bernhard, 137, 385, 387, 395, 451,
 456–458, 471, 474–476, 478, 479, 489,
 492, 504, 506–507, 518, 526, 532

 Ries, Adam, 328
 Riesz, Frigyes, 412, 527
 Robert of Chester, 318, 324
 Robertson, E. F., 196
 de Roberval, Gilles Personne, 365, 366,
 372, 382
 Robins, Gay, 58, 71, 74
 Robinson, Abraham, 399
 Robson, Eleanor, 29, 46, 48
 Roch, Gustav, 506
 Rogers, Douglas, 54
 Rolle, Michel, 379, 381
 Rosen, Frederic, 294, 297
 Rothman, T., 278
 Rudin, Walter, 495
 Ruffini, Paolo, 4, 5, 258, 440–442
 ibn Rushd, 291
 Russell, Bertrand, 291, 372, 375, 540, 549

 Sabra, A. I., 305
 Saccheri, Giovanni, 303, 482–484, 493
 ibn Sahl, Abu Saad, 306, 393
 Salmon, George, 473
 Santayana, George, 3, 11
 Sargon, 28
 Sarkor, Ramatosh, 206
 Sawaguchi Kazuyuki, 271, 273
 Scharlau, Winfried, 524
 Schwarz, Hermann Amandus, 533
 Schweikart, Ferdinand Karl, 485, 487
 “Scorpion”, 56
 Scott, Charlotte, Angas, 409
 von Seidel, Philipp Ludwig, 525
 Seki Kōwa, 267, 269–270, 272, 280, 453
 Seki Takakazu, 267
 Seleucus, 29
 Seneta, E., 429
 Sentlowitz, Michael, 290
 Senusret I, 57
 Serret, Joseph, 477
 Sesiano, Jacques, 98, 99
 Sesostris, 68
 Shakespeare, 81
 Shannon, Claude, 11, 31
 Shen Kangsheng, 264
 Shidhara, 209
 Shih Huang-Ti, 240
 Shimura, Gorō, 270
 Shute, Charles, 58, 71, 74
 Siegmund-Schultze, Reinhard, 508
 Sigler, L. E., 326, 327
 Simmias, 86

- Simms, D. L., 149
 Simplicius, 81, 116, 117, 131
 Simpson, Edward Hugh, 431
 ibn Sina, 291
 Skinner, Burrhus Frederic, 20, 21, 427
 Smith, D. E., 31, 126, 272, 273, 279, 361, 498
 Snell, Willebrod, 306, 393
 Socrates, 47, 86, 115, 129, 135
 Socrates Scholasticus, 197
 Sopatros, 129
 Sophocles, 117
 Sotion, 130
 Spengler, Oswald, 414
 Sporos, 123
 Srinivasiengar, C. N., 205, 218
 Stäckel, Paul, 472, 489
 Stalin, Joseph, 414
 Steiner, Jacob, 199, 451, 463
 Stephanus, 82
 Stevin, Simon, 328, 340, 434
 Stirling, James, 423
 Stobaeus, 116
 Strabo, 177
 Struik, Dirk, 355, 434, 466, 473
 Struve, Friedrich Wilhelm (Vasilii Yakovlevich), 488
 Struve, Vasilii Vasil'evich, 72
 Sturm, Charles, 515
 Suiseth, Richard, 331
 Sun King, 244
 Sun Zi, 242, 246, 254
 Suslin, Mikhail Yakovlevich, 535
 Swetz, Frank, 243, 250
 Swineshead, 331
 Swyneshed, Richard, 331
 Sylvester II, 316
 Synesius, 197

 Takebe Kenkō, 269–270, 275, 277, 280
 Taliaferro, R. Catesby, 160
 Tannery, Jules, 539
 Tannery, Paul, 98
 Tarik, 283
 Tarski, Alfred, 540
 Tartaglia, Niccolò, 321, 338
 Taurinus, Franz Adolph, 486
 Taussky, Olga, 414
 Taylor, Brook, 379, 390, 464, 512
 Taylor, Richard, 100
 Thales, 79–82, 84, 103, 123
 Theaetetus, 83, 129
 Theano, 197

 Theatetus, 110
 Theodorus, 83, 110
 Theodosius, 190, 196
 Theomedus, 130
 Theon of Alexandria, 78, 81, 89, 140, 169, 171, 172, 190, 191, 196–199, 313
 Theon of Smyrna, 81, 117
 Theseus, 135
 Thomas, Ivor, 89
 Thoreau, Henry, 24
 Thoren, Victor E., 334
 Thoth, 62
 Thureau-Dangin, François, 42
 Thutmose IV, 68
 Thymaridas, 98
 Timur the Lame, 204, 287
 Todhunter, Isaac, 392, 418
 Toomer, G. J., 160, 286
 Torday, Emil, 17
 Torricelli, Evangelista, 365
 Tsaban, Boaz, 70
 Tschirnhaus, Ehrenfried, 434–436, 439
 Turán, Paul, 244
 al-Din al-Tusi, Sharaf al-Din, 301
 al-Tusi, Nasir al-Din, 289, 308, 309, 331
 al-Tusi, Sharaf, 258, 289, 363
 al-Tusi, Sharaf al-Din, 299, 300
 Tzetzes, Johannes, 129, 149

 Ulugh Beg, 287
 Unguru, Sabetai, 105, 160, 161, 165
 al-Uqlidisi, 292
 Ursus, 334

 Valéry, Paul, 362
 Varadarajan, V. S., 210
 Venn, John, 425
 Viète, François, 322, 341, 342, 346, 359, 376
 da Vinci, Leonardo, 345, 349
 Vitruvius, 80, 149
 Vivanti, Giulio, 537
 Vogel, Kurt, 71, 329
 Voils, D. L., 54
 Volkov, Aleksei, 264
 Voltaire, 375
 Volterra, Vito, 526

 van der Waerden, Bartel Leendert, 19, 41, 42, 46, 59, 67, 72, 193, 413
 Wallace, William, 381
 Wallis, John, 370, 497–499
 Wang Lian-tung, 267

- Wang Pu-son, 267
Wang Xiaotong, 257, 260
Wantzel, Laurent, 444
Waring, Edward, 441, 449
Watson, G. N., 211
Weber, Ernst Heinrich, 425
Weber, Wilhelm, 477
Weierstrass, Karl, 387, 389, 391, 395–398, 400, 406, 479, 504, 507–508, 526, 527, 537
Weil, André, 230
Werner, Johann, 334
Wessel, Caspar, 498
Weyl, Hermann, 413, 477
Whitehead, Alfred North, 550
Whiteside, Thomas, 362, 448, 521
Whittaker, Edmund Taylor, 11
Wigner, Eugene, 9
Wiles, Andrew, 100
Winston, Mary Frances, 409
Woepcke, Franz, 298
Woodhouse, Robert, 392, 394
Wylie, Alexander, 247
Xu Guangchi, 243
Yackel, Carolyn, 18
Yang Hui, 243
Yoshida Koyu, 269, 273
Young, Grace Chisholm, 408–411, 414
Young, William Henry, 409
Yule, George Udny, 431
Zeno, 107–108, 113, 139, 375, 543
Zenodorus, 78, 88, 169–171, 176, 186, 191
Zermelo, Ernst, 537, 549, 551
Zeuthen, H. G., 341
Zhao Shuang, 241, 250, 251, 253
Zhao Youqin, 264
Zhen Luan, 255
Zhmud, Leonid, 81
Zhu Shijie, 269
Zu Chongzhi, 202, 243, 264–266, 274
Zu Geng, 202, 243, 264
Zverkina, G. A., 92

SUBJECT INDEX

- abacus, 268, 314, 316, 323
- Abbasid Dynasty, 283, 285
- Abel–Poisson convergence, 517
- abelian group, 443
- abelian integral, 407, 444
- abscissa, 164, 367, 380, 389
- absolute, 456
- absolute continuity, 528
- Academy, 84, 86, 115, 128, 129, 134, 171
- Academy of Sciences
 - Belgian, 553
 - Berlin, 437
 - Danish, 469
 - Indiana, 70
 - Paris, 381, 387, 423, 444, 459, 476, 478, 505, 538
 - Russian, 376, 386, 424, 438, 513
- Accademia dei Lincei, 479
- accounting, 5
- Achilles paradox, 108
- acre, 30
- Acta eruditorum*, 377, 388, 437
- actinium, 430
- acute angle, 332
- adding machine, 345
- addition, 12, 38, 59, 245
- addition formula, 502, 503
- administration, 30
- Aegean Sea, 80
- affine, 452
- Afghanistan, 289
- Afghans, 204
- Africa, 28, 179, 283
- Agisymba, 179
- agrimensor*, 174
- aha* computation, 64
- Ainu, 31
- air, 84
- Akhmim Wooden Tablet, 57, 62
- Akkadian civilization, 27
- Akkadian language, 29, 39
- akoustikoi*, 84
- al-jabr*, 287, 294
- alchemy, 24, 285
- alcohol, 76, 285
- Aldebaran, 285
- alea*, 418
- aleatics*, 418
- \aleph_0 , 535
- Aleppo, 289
- Alexander of Macedon, 206
- Alexandria, 20, 57, 68, 72, 77, 83, 87, 116, 123, 140–147, 149, 154, 161, 172, 177, 184, 190, 191, 196, 197, 199, 206, 286
- Algebra*, 233, 296–300, 318, 324
- algebra, xxv, 4–5, 8, 13, 21, 44, 46, 66–77, 91, 98, 100, 106, 192, 201, 205, 218, 228–235, 239, 242, 246–248, 250, 255–266, 270–277, 287, 289, 292–301, 320, 324, 340, 372, 373, 386, 433–447, 477, 481, 543
 - “syncopated”, 99
 - computer, 1, 6
 - fundamental theorem, 544
 - geometric, 105, 113
 - linear, 247
 - modern, 4
 - multilinear, 474
 - symbolic, 99
- algebraic curve, 293
- algebraic extension, 443
- algebraic function, 387
- algebraic geometry, 492
- algebraic number, 295, 523, 534
- algebraic operation, 5
- Algeria, 319

- Algol, 285
 algorism, 294
 algorithm, 184, 287, 294
 Euclidean, 92–93, 101, 110, 144, 146, 231, 245, 264
 mutual-subtraction, 264
 Alhazen's problem, 305
 allegory of the cave, 129
Almagest, 77, 81, 89, 161, 177, 178, 182, 184–187, 196, 286, 308, 318, 320
 almanac, 285
 almost everywhere, 422
 almost surely, 422
 Altair, 285
 altar, 205, 213, 214
 alternating, 272
 altimetry, 317
 altitude, 112, 131, 138, 172, 194, 215, 219, 225, 252, 332, 337
 Alzheimer's disease, 418
amblytomē, 119
 amicable numbers, 293
 Amorite civilization, 28
 amplitude, 10
 analog computer, 344
Analyse des infinement petits, 379, 382
 analysis, 192, 371, 386, 450, 468, 532
 complex, 463, 495–510
 functional, 530, 537
 nonstandard, 376, 399
 real, 521–531
Analysis situs, 459–461
 analytic function, 389, 500–508, 511, 521
 analytic geometry, 288, 330, 337, 358–362, 373, 492, 521
Analytic Geometry of Three Dimensions, 473
 analytic set, 535
Ancient and Modern Mathematics, 271
 Andaman, 31
 angle, 30, 48, 53, 136, 141, 145, 173, 184, 187, 195, 199, 220, 225, 303, 332
 acute, 85, 125, 332
 base, 82, 279
 central, 143, 184
 dihedral, 456
 exterior, 137
 incidence, 306, 393
 inscribed, 82, 143
 obtuse, 85, 308
 refraction, 306, 393
 right, 82, 85, 122, 146, 168, 185, 220, 308
 summit, 482
 trisection, 106, 116, 122–125, 139, 147, 286, 297, 305, 342, 444
 anharmonic ratio, 351
 animal behavior, 14
 animals
 mathematical reasoning, 20–21
Annales des sciences naturelles, 325
 annulus, 151
 antidifferentiation, 397
 antilogarithm, 346
 antipodal points, 395
 Antiquités Orientales, 39
 antisymmetrizing, 474
Anuyoga Dwara Sutra, 206
 AO 6484, 39, 44
 AO 6670, 42
 AO 8862, 41
Apastambe Sutra, 216
 apex, 120, 162
 aphelion, 183
 apogee, 182
 Apollo, 117, 135
 apothem, 144
 application, 103–105, 114, 162, 165
 with defect, 104, 113, 142, 146, 161, 162
 with excess, 104, 142, 145, 161, 162
 applied mathematics, 8
 approximation, 337
 aqueduct, 176
 Arabia, 204
 Arabic language, xxvi, 20, 79, 81, 88, 89, 160, 161, 204, 283–286, 292, 305, 317, 323, 327, 482
 Arbela, 178
 arc, 48, 143, 179, 180, 184, 193, 195, 220, 224, 236, 271, 275, 332, 368
 arc length, 388, 464
 archaeology, 14
 archery, 242
 Archimedes' principle, 132, 144, 194, 303, 376
 Archimedes' tomb, 148–149
 architecture, 80, 203, 290, 314
 arcsine, 388
 Arctic Circle, 178
 area, xxv, 6, 19, 30, 41, 43, 46, 48, 55, 58, 69–73, 80, 145, 150, 169, 172, 176, 195, 198, 214, 219, 238, 242, 247, 251, 271, 320, 328, 330, 374, 484, 526
 “practical”, 227
 “rough”, 227
 transformation, 114, 215

- areas
 application, 103–105, 114
*Arithmetic of Boethius*⁷, 316
 arithmetic, 5, 8, 13, 21, 38, 56–66, 86, 98, 129,
 201, 205, 210, 213, 241, 242, 287, 288,
 313, 320, 323, 349
 commercial, 320
 fundamental theorem, 97
*Arithmetic Classic of the Gnomon and the
 Circular Paths of Heaven*, 241
Arithmetic in Nine Chapters, 258
 arithmetic operations, 12
 arithmetic progression, 229
Arithmetica of Diophantus, 97–101, 198, 219,
 295, 327
Arithmetica of Nicomachus, 92, 93
Arithmetica of Diophantus, 81
Arithmetica universalis, 436
Arithmetical Classic of the Zhou, 201
 arithmology, 418, 529
 Arpinum, 148
 arrow paradox, 108, 375
 arrows, 233
Ars conjectandi, 421–422, 431
Ars magna, 322, 339
 art, 290
Artis analyticae praxis, 322, 341
Aryabhatiya, 207, 219, 222, 224
 Aryan civilization, 203
 Asia, 179, 203
 Asia Minor, 80
 Askelon, 81
 associative operation, 474
 assumption, 90
 Assyria, 28
 Assyrian Empire, 28
 asterism, 208
 astral geometry, 485
 astrolabe, 285, 286, 305, 317
 astrology, 19, 24, 28, 79, 206, 241
 astronomy, 5, 48, 78, 79, 82, 83, 129, 177–189,
 208–210, 220, 221, 230, 239, 241, 244,
 264, 284, 285, 288, 292, 308, 313, 315,
 317, 320, 333, 410, 476, 488
 geocentric, 181, 189, 318
 heliocentric, 182, 189
 asymptote, 161, 164, 171, 298
 Athenian Empire, 28, 79, 117
 Athens, 77, 84, 115–139
 Atsuta, 277
Attic Nights, 84
 attribute, 135
 Aurillac, 316
Ausdehnungslehre, 474
 Australia, 31
Auswahlprinzip, 538
 average, 39–41, 44, 50, 55, 142, 228, 252
 average speed, 196
 axial triangle, 119, 162
 axiom, 7, 13, 82, 244
 Playfair's, 171
 Zermelo's, 537
 axiom of choice, 9, 536–540, 554
 axiomatic method, 98
 axis, 161, 163, 193, 253
 Azerbaijan, 290
 Babylon, 28, 184
 Babylonian civilization, 68, 79
 Babylonian language, 29
 Baghdad, 57, 204, 284, 286–289,
 302, 308
 Baire category theorem, 530, 531, 540
Bait al-Hikma, 286
 Bakshali manuscript, 20, 206, 218, 219
 ballistics, 498
 bamboo, 7
 Banach–Tarski paradox, 540, 541
 band-limited signal, 10
 banker, 421
 barycentric calculus, 452
 barycentric coordinates, 452
 base, 131, 138, 170, 191, 194, 215,
 252, 265
 of a pyramid, 49, 55, 74
 of a triangle, 112
 of logarithms, 31
 of numeration, 31–33, 37
 base angle, 82
 basket, 73
 Bath, 318
 Battle of Arbela, 116, 178
 Battle of Constantinople, 283
 Battle of Marathon, 115
 Battle of Plataea, 115
 Battle of Thermopylae, 115
 beer, 67, 76
 bees, 16, 191
 begging the question, 136
 Beijing Library, 244
 bell-shaped curve, 428, 430
 Bengali language, 203
 Bergama, 88
 Berkeley, 379, 431

- Berlin, 43, 72, 386, 406, 412
 Berlin papyrus 6619, 67, 72
 Bernoulli number, 422
 Bernoulli trials, 421–423
 Betelgeuse, 285
 Betti number, 478
Bhagabati Sutra, 206, 217
bhuja, 220
 Bianchi identity, 479
 bias, 431
 Bible, 28, 70
Bibliotheca mathematica, 537
 Bibliothèque Nationale, 320
 bicylinder, 263
 Bijapur, 209
 binary operations, 39
 binary system, 31
 binomial, 228
 binomial distribution, 428
 binomial series, 517
 binomial theorem, 371, 373, 378, 421, 506
 birds, 14
 bisection, 276
 bisector, 146, 170, 173, 180, 187, 199, 215, 309
 Bisutun, Iran, 29
 bit (binary digit), 31
 Black Sea, 27
 Blessed Islands, 179
 block printing, 239
 BM 13901, 46, 55
 BM 85194, 49, 228
 BM 85196, 47, 54
 Bobbio, 316
Bodhayana Sutra, 213, 215, 216
 Bodleian Library, 294
 Bologna, 321, 413
 Bolzano–Weierstrass theorem, 533
 Bombay (Mumbai), 210
Book of Changes, 255
Book of Completion and Reduction, 287
Book of Lemmas, 149
Book of Squares, 326
Book on the Resolution of Doubts, 306
Book on Unknown Arcs of a Sphere, 308
 Boolean ring, 548
 Borel sets, 535
 boson, 427
 Boston Museum of Fine Arts, 57
 botany, 239
 boundary, 511
 bowstring, 221, 249, 305, 318
 Boxer Rebellion, 240
 brachistochrone, 392, 469
 “Brahmagupta’s identity”, 230
 Brahmagupta’s theorem, 227
 Brahman, 208
Brahmasphutasiddhanta, 209, 227–229, 285
 branch, 352
 branch point, 458, 506
 bread, 58, 67
 Brescia, 321
 Brianchon’s theorem, 449
 bricks, 213
 Britain, 178, 315
 British Museum, 38, 57
 Brooklyn Museum of Art, 57
 Brouwer fixed-point theorem, 556
 Bryn Mawr College, 405, 409, 413, 414
bu, 253
 buckminsterfullerene, 350
 buckyball, 351
 Buddha, 409
 Buddhism, 202, 204, 206, 267, 269, 277, 290
 Burali-Forti paradox, 539, 549
 Bushoong, 17
 buyer, 425
 Byzantine Empire, 81, 283, 291, 313

 c, 535
 Cabo Delgado, 179
 Cairo, 58, 141
 Caius College, 425
 calculating devices, 267–268
 calculating machines, 345
 calculator, 30, 39, 184, 329
 calculus, 8, 157, 195, 201, 205, 258, 274–277, 311, 358–401, 406, 448, 464, 468, 491, 500, 502
 barycentric, 452
 foundations, 383, 397–399
 fundamental theorem, 370, 377, 526
 integral, 235, 377
 priority dispute, 381–382
 calculus of residues, 505
 calculus of variations, 312, 374, 379, 387, 391–397, 400
 Calcutta, 210, 411
 calendar, 12, 80, 150, 230, 241, 244, 320
 Gregorian, 373
 proleptic, 373
 Julian, 373, 406
 lunar, 285
 lunisolar, 241, 285
 Muslim, 285

- California, 379
 California State University, 466
 Caliphate of Cordoba, 283, 308
 calligraphy, 242
 Cambridge University, 100, 370, 373, 408, 425
 Cambridge University Press, 411
 camel, 219
 canal, 68, 248, 250, 287
 Canary Islands, 179
 cancelation, 92, 234
Candide, 375
 Cantor–Bendixson theorem, 535
 Cardano formula, 338, 436
 cardinal number, 534
 cardinality, 535
 Carolingian Empire, 313, 315
 carpentry, 220
 carrying, 38
 Cartesian product, 9
 Carthage, 148
 carts, 246, 253
 casino, 9
 Caspian Sea, 27
 Catch-22, 415
 categoricity, 544
 cathedral school, 315, 318
 Cauchy convergence criterion, 525
 Cauchy integral formula, 505, 516
 Cauchy integral theorem, 505
 Cauchy sequence, 525
 Cauchy–Riemann equations, 505, 520
 Cavalieri’s principle, 154, 243, 263, 274, 365–384
 Cayley metric, 456
 “celestial element method”, 271
 celestial equator, 224
 celestial sphere, 177, 180, 224, 317, 479
censo, 320, 335
 center, 163, 215, 348
 center of curvature, 466, 479
 center of gravity, 193
 central limit theorem, 424, 428–429
Centrobaryca, 195
 centroid, 195, 198, 262
centuria, 174
Ceyuan Haijing, 271
 chain, 459
 chain of being, 204
 Chaldean (astrologer), 84
 Chaldean Empire, 28
 chance, 427
Chandahsutra, 205
 change of variable
 fractional-linear, 45
 linear, 41
 charioteership, 242
 Chebyshev’s inequality, 428, 430
 chemistry, 239, 373
cheng, 245
chi, 251, 252
 Chicago, 290, 410
 children, 15
Chiliades, 129
 China, 6, 22, 30, 37, 46, 179, 201, 204, 206, 218, 219, 239–267, 284, 290, 300, 308, 310, 338, 447
 number system, 32
 Chinese language, xxvi, 203, 221, 241, 267
 Chinese remainder theorem, 224, 246, 294
 Chinese writing, 267
 Ching (Manchu) Dynasty, 240
chong cha, 251
ch’onwonsul, 271
 chord, 152, 165, 166, 177, 180, 184, 220, 276, 304, 333, 490
 of half-angle, 186
 chords
 table, 184–187
 Christians, 197
CHŪ, 267
 circle, xxv, 6, 16, 30, 33, 48, 68, 70–71, 79, 106, 122, 123, 131, 136, 143, 151, 160, 162, 165, 168, 169, 177, 180, 181, 183, 211, 216, 251, 262, 348, 350, 359, 371, 462
 circumscribed, 48, 173, 216
 equal to a given square, 215, 228
 equatorial, 224
 generating, 366
 great, 219, 235
 inscribed, 173, 216, 263, 278
 measurement, 150
 osculating, 464, 467
 quadrature, 106, 115–117, 131, 139, 147, 153, 215, 219, 323, 377, 445
 Egyptian approximation, 70
 rectification, 275, 277
 circumference, 48, 169, 184, 216, 219, 251, 261, 279, 372
 “practical”, 227
 citizenship, 248
citra kardinem, 174
 Civil War, English, 373
 classical problems, 127
 clay tablets, 27

- clock, 31, 178, 393
 cloth, 336
 Cloyne, 379
 cluster point, 533
 Cnidus, 129, 130
 co-declination, 224
 co-latitude, 224, 235
 coefficient, 266, 433
 coin, 206
 coin-tossing, 424
Collection, 140, 161, 172, 190–196, 199
 collinearity, 453
 color, 234
 Columbia University, 49
 combination, 233, 421
Combination Book, 275
 combinatorial coefficients, 217
 combinatorics, 205, 217–218, 233–234, 239
 combinatory product, 474
 comet, 426
Commentaries of Pythagoras, 197
Commentarii, 424, 438
*Commentary on the Premises to Euclid's Book
 The Elements*, 306
 commentator, 80–82, 90, 102, 153,
 190–199
 commerce, 30, 58, 79, 80, 287, 296
 common divisor, 92
 Communism, 240
 commutative, 548
 compass, 186, 191, 242, 305, 444, 484, 522
Compendium, 377
 completeness, 544, 551, 554
 completing the square, 235, 295
 complex analysis, 8, 387, 463, 495–510
 complex number, 340, 346, 379, 389, 445, 448,
 453, 495–496, 513
 complex plane, 449, 452
 complex variable, 401, 500
 component, 458
 composite number, 93
 composite ratio, 133, 138, 145, 193, 194
 composition, 444
 compound interest, 44
Comptes rendus, 387, 538
 computation, 6, 209
 Egyptian, xxiv, xxv
 computer, 6, 18, 225
 computer algebra, 1, 6, 45
 concavity, 467
 conchoid, 123, 125–169
 conditional probability, 23, 419
 conditioning, 16
 cone, xxv, 118, 131, 144, 169, 352, 472
 acute-angled, 161
 nappe, 164
 obtuse-angled, 161
 right-angled, 161
 slant height, 152
 cones, similar, 144
 conformal mapping, 464, 469, 473
 Confucianism, 239
Congrès scientifique de France, 325
 congruence, 201, 492
 congruent, 235
 congruent triangles, 21
 conic section, xxv, 88, 103, 118, 123, 160–166,
 169, 193, 298, 308, 353, 359, 377, 449,
 453, 454
 subcontrary, 162
 conical projection, 180
Conics, 140, 192, 286, 305, 308
 conjugate points, 395
 conjunction, 547
 connectedness, 461
 connectivity, 458, 478
 conoid, 150
Conoids and Spheroids, 150
 conservation law, 387
 conservation of energy, 476, 479
 consistency, 544, 551, 554
 constant, 347, 359
 Euler's, 522
 constant of proportionality, 163
 Constantinople, 82, 153, 197, 284, 285
 constrained extremum, 394
 constructivism, 536
Continuation of Ancient Mathematics, 257
 continued fraction, 101
 continuity, 8, 400, 461, 492, 495, 499, 522
 absolute, 528
 continuous medium, 6
 continuous quantity, 5, 6
 continuum, 107, 108, 129, 462
 continuum hypothesis, 535, 540
 convergence
 Abel–Poisson, 517
 in measure, 422
 in probability, 422
 pointwise, 525–526
 uniform, 525–526
 convergence factor, 517
 converging lines, 307
 convexity, 470

- coordinate
 - radial, 181
- coordinate lines, 361
- coordinate system, 17
- coordinates, 174, 311
 - barycentric, 452
 - Cartesian, 359
 - homogeneous, 454
 - line, 455
 - point, 455
 - polar, 368, 518
 - rectangular, 368, 518
 - spherical, 181
- Copenhagen, 498
- Cordoba, 302
- Cornell University, 409
- corner (Egyptian square root), 67
- corner condition, 396
- Corpus Agrimensorum Romanorum*, 174
- cosa*, 20, 320, 335
- cosecant, 221
- coset, 444
- cosine, 188, 221, 224, 225, 237, 275, 335, 337, 342, 358, 501
 - hyperbolic, 484
- cosmimetry, 317
- cotangent, 221
- countable set, 538
- counting, 14–16, 80
- counting board, 18, 38, 245, 256, 268, 274
- counting rods, 18, 245, 268
- Cours d'analyse*, 514, 525
- cousinhood, 7
- covariant, 455
- crafts, 16
- Cramér's paradox, 453, 463
- Cramér's rule, 453
- credit, 336
- Crest Jewel of the Siddhantas*, 235
- Crete, 118
- cross product, 446, 474
- cross-ratio, 351, 491
- crosscut, 458
- Croton, 84, 128, 197
- Crusades, 284
- crystal ball, 19
- crystallography, 325
- cube, 38, 96, 97, 106, 117, 130, 144, 218, 228, 254, 264
 - doubling, 106, 107, 116–122, 139, 147, 342, 444
- cube root, 38, 229, 257, 347, 496
- cubic curve, 449
- cubic equation, 40, 234, 257, 261, 265, 286, 289, 296, 298, 308, 321, 338–340, 346, 433, 438, 496
 - irreducible case, 340, 342–343
 - two-variable, 462
- cubic polynomial, 407
- cubit, 69, 70, 73
- cun*, 252
- cuneiform, 19, 22, 27, 29, 31, 36, 38, 47–49, 54, 58, 65, 82, 83, 109, 142, 228
- current, 477
- curvature, 251, 374, 464, 466, 477, 489
 - center, 479
 - Gaussian, 470, 471, 479
 - geodesic, 473
 - normal, 512
 - radius, 464, 468, 479, 483, 487
- curvature tensor, 479
- curve, 193, 205, 360
 - algebraic, 293, 453
 - bell-shaped, 399
 - length, 475
 - plane, 162, 464–468
 - space, 464, 469
 - space-filling, xxiii
 - transcendental, 453
- curves, homologous, 459
- curvilinear problem, 123
- Cutting Off of a Ratio*, 161
- cycle, 460
- cyclic quadrilateral, 173, 227
- cycloid, 364–365, 372, 374, 393, 453, 465
 - area, 366
 - curtate, 453
 - prolate, 453
 - tangent to, 364
- cycloidal pendulum, 465, 467
- Cyclops, 94
- cyclotomic equation, 443
- cylinder, 80, 118, 131, 144, 148, 150, 154–155, 191, 195, 253, 262, 352, 472
 - area, 72
- cylinders, similar, 144
- Cyzicus, 130
- Daniel (book of the Bible), 28
- dār al-'ilm*, 286
- Dark Ages, 291
- Dasagitika Sutra*, 208
- Data*, 87, 140, 144, 192, 286, 307, 330
- day, 12, 33

- day-circle, 224
 daylight, 178
De arte combinatoria, 421
De configurationibus qualitatum motuum, 331
De divina proportione, 349
De institutione musica, 315
 De Morgan's laws, 545
De motu stellarum, 318
De numeris datis, 329
De pictura, 321
De quadratura arithmetica circuli, 522
De ratiociniis in ludo aleæ, 420
De revolutionibus, 309
De Thiende, 328
De triangulis omnimodis, 309, 320, 331
 decagon, 185
 decimal expansion, 12, 522
 infinite, 329
 decimal system, 31, 37, 59, 188, 207, 212,
 213, 292
 declination, 177, 180, 224
Decline and Fall of the Roman Empire, 198
 decorum, 242
decumanus maximus, 174
 Dedekind cut, 525
Dedoména, 144
 defect, 104, 163, 485
 deferent, 181, 188
 deficient number, 94
 deficit, 352
 definition, 82, 90, 244
 degree, 33, 49, 79, 184, 211, 222
Della pittura, 321, 352
 Delos, 117, 130
 demonstration, 6
 denarius, 206
 density, 133, 511
 dependent trials, 429
 dependent variable, 514
 derivative, 299, 358, 363, 370, 373, 374, 376,
 382, 384, 464, 512
 partial, 520
 second, 464
 derived set, 533, 535
 Desargues' theorem, 192, 354
 descriptive set theory, 535
*Deatiled Analysis of the Mathematical Rules
 in the Jiu Zhang Suan Shu*, 243
 determinant, 272, 453
 determinate problem, 99
 determinism, 432
 developing a function, 548
dextra decumani, 174
 diagonal, 47, 109, 113, 122, 130, 144, 146, 184,
 214, 216, 249, 273, 309, 384
 diameter, 48, 71, 117, 123, 132, 144, 152, 161,
 185, 216, 219, 227, 237, 252, 254, 261,
 275, 279, 309
 dice, 419, 430
 Dichotomy paradox, 107
Dictionary of Scientific Biography, 289
 differentiability, 411
 differentiable manifold, 471
 differential, 376, 378, 394
 differential equation, 312, 379, 382, 383, 387,
 401, 407, 459, 460, 468, 473, 495
 exact, 387
 ordinary, 387–389
 partial, 390–391, 425
 normal form, 391
 differential form, 478
 differential geometry, 486, 492
 differentiation, 382, 500
 difformly difform, 330
 digit, 234
 dimension, 41, 69, 96, 144, 193, 337, 359,
 361, 535
 fourth, 308
 invariance, 556
dīnāra, 206
 Diophantine equation, 99, 213, 229, 230,
 254
dirhem, 295
 Dirichlet function, 515
 Dirichlet's principle, 395
 discontinuity, 529
Discourse on Method, 360
Discourses on the Seven Sages, 82
 discrete, 107
 discriminant, 342
 discrimination, 431
Discussion of difficulties in Euclid, 307
 disjunction, 547
 disk, 68, 179, 255
 equatorial, 235
 dispersion, 425
Disquisitiones arithmeticae, 247
*Disquisitiones generales circa superficies
 curvas*, 469
 dissection, 201, 262, 264
 distance, 253, 320
 distribution, 512
 binomial, 428
 Gaussian, 427

- normal, 427, 428, 430
 Poisson, 428
 dividend, 34, 329
 divination, 19, 24, 241, 255, 427
Divine Comedy, 317
 Divine Proportion, 349
 divisibility, 97, 246
 infinite, 108
 division, 12, 19, 34, 38, 59, 101, 199, 245, 322
 Egyptian, 61
 polynomial, 445
 divisor, 329
 common, 92
Doctrine of Chances, 423
 dodecahedron, 144
 dogs, 20
 perception of shape, 16
 dominated convergence theorem, 527, 530
 Doric, 118
 dot product, 446
 double difference, 251
 double square umbrella, 263
 doubling, 59
 doubling a square, 47, 214
 doubling the cube, 106, 107, 116–122, 130,
 139, 147, 342, 444
 doubly periodic function, 504
 dozen, 30
 drachma, 206
 dram, 30
dramma, 206
 duality, 449, 452
 Duquesne University, 494
 Dutch, 270
 dyad, 72

e, 12, 31, 423, 425, 445, 522
 earth, 84, 96, 177–180, 188, 208, 211, 220, 251,
 289, 317, 488
 earth–moon system, 182
 Easter, 19
 eccentric, 181
 eccentricity, 418
Ecclesiastical History, 197
 eclipse, 181, 190, 238, 244
 lunar, 82
 solar, 82
 ecliptic, 180, 183
 Ecole Normale, 444
 Ecole Polytechnique, 398, 450
 economics, 425
 ecu, 336

 edge law, 459
 Edict of Nantes, 423
 education, 209, 210
 effective enumeration, 534
 Egypt, 18, 19, 22, 25, 28–30, 32, 37, 56–77, 79,
 80, 83, 84, 90, 98, 116, 130, 138, 140,
 141, 179, 205, 235, 244, 283, 317,
 319, 376
 Lower, 56
 Upper, 56
 Egyptian computation, xxiv, xxv
 Egyptian Museum, 58
 Egyptian numeration, 95
eida (species), 98
 Eighteenth Dynasty, 68
 Elamite language, 29
 elasticity, 394, 473, 476
 electricity, 477
 element, 84
Elements, 7, 77, 81, 86, 91, 94, 97, 103, 112,
 116, 117, 130, 132, 136, 140–147, 152,
 158, 165, 169, 171, 176, 185, 194, 196,
 198, 210, 214, 243, 286, 288, 290, 302,
 307, 318, 327, 348, 373
 ellipse, 16, 17, 118, 161, 165, 183, 348, 352,
 372, 455, 462, 501
 definition, 162
 eccentricity, 16
 equation, 163
 rectification, 270
 string property, 166, 168
 elliptic function, 97, 386
 elliptic integral, 407, 445, 502
emolumentum, 425
 Emperor Yu, 255
 empirical knowledge, 360
 empty place, 207, 213
Encyclopédie, 386, 397
 energy
 conservation, 476
 engineering, 53, 56, 88, 174–175, 322, 329,
 387, 448
 England, 211, 318, 411, 423
 English language, 28, 31, 39, 59, 160, 163, 204,
 217, 241, 247, 351, 374, 412, 417,
 485, 538
 entropy, 428
 enumerable number, 217
 envelope, 468, 469
epanthēma, 98
 Ephesus, 85
 epicycle, 181, 188

- epistemology, 8, 85
 equalation, 352
 equality
 exact, 129
 equation, 40, 44, 76, 163, 205, 218, 265, 287, 401, 433–447
 approximate solution, 255
 cubic, xxv, 40, 234, 257, 265, 286, 289, 296, 298, 308, 321, 338–340, 346, 433, 438, 496
 irreducible case, 340, 342–343
 two-variable, 462
 cyclotomic, 443
 differential, 312, 382, 387, 401
 Diophantine, 99, 229, 230, 254
 Euler's, 394, 395
 gravitational, 479
 heat, 390, 399, 519
 Laplace's, 513, 519
 linear, 40, 219, 230, 242, 272, 296
 linear system, 256
 numerical solution, xxiv, 258, 266
 Pell's, 229, 237
 greater solution, 230
 lesser solution, 230
 quadratic, xxv, 19, 40, 69, 99, 105, 228, 233, 256, 272, 296, 301, 329, 338, 359, 433, 447
 quartic, 234, 260, 308, 322, 338–340, 346, 433, 438
 quintic, 4, 439, 442, 445
 solution by radicals, 338
 wave, 390, 512, 513, 519
 equations
 Cauchy–Riemann, 505, 520
 Frenet–Serret, 477
 Mainardi–Codazzi, 477
 Navier–Stokes, 473
 Equator, 179, 235
 equatorial circle, 224
 equidistant curve, 307, 309
 equilateral number, 110
 equilateral triangle, 279, 316
Equilibrium of Planes, 149
 equinox, 224
 vernal, 180
 Erbil, Iraq, 178
 Erlangen, 412
Essai sur une manière de représenter les quantités imaginaires dans les constructions géométriques, 499
 Ethiopia, 179
 Euclidean algorithm, 92–93, 101, 110, 144, 146, 231, 245, 264, 445
 Euclidean geometry, xxv, 6, 348, 367, 372, 376, 480
Euclides ab omni nœvo vindicatus, 482
 Euler characteristic, 456, 461, 473
 Euler constant, 523
 Euler's equation, 394, 395
 Euler's formula, 459
 Euphrates River, 27, 28, 288
 Eureka College, 410
 Europe, 81, 96, 179, 187, 247, 266, 267, 270, 284, 311–338, 358, 406, 482
 European Union, 30
 even number, 72, 93, 110, 255
 evenly even number, 93
 event, 23, 419, 547
 evolute, 464, 467, 479
 evolution, 24
 exceedence, 352
 excess, 104
 excluded middle, 9
 existence, 8, 145, 529, 534, 544, 546
 expectation, 420, 428
 exponent, 44, 97, 218, 336
 exponential function, 347, 378, 500, 545
 exterior-angle principle, 137
 extremal
 strong, 397
 weak, 395, 397
 extremes, 143
 face, 17
 faction, 434
 fairness, 248, 301
Fakhri, 327
 falconry, 319
 false position, 66
 fathom, 30
fen, 251, 253
 Fermat's last theorem, 89, 91, 100, 461
 Fermat's principle, 392, 394
 fermion, 427
 Fertile Crescent, 29
Fibonacci Quarterly, 325
 Fibonacci sequence, 324–326
 field, 4, 247, 443, 536
 algebraically closed, 433, 447
 non-Archimedean, 383
 ordered, 383
 Fields Medal, 461
 figurate number, 77, 91, 95–118

- figure, 17
 finger reckoning, 292
 finite difference, 379
 fire, 84, 96
 first category, 529
 First Crusade, 284
 first fundamental form, 471, 477, 479
 first ratio, 397
 five-line locus, 193
 flat manifold, 476
 flat surface, 468
 flavor, 217
Floating Bodies, 150
 floating-point number, 6
 floating-point system, 32
 Florida, 54
Flos, 327
 fluent, 374, 382
 fluxion, 373–374, 379, 381, 382
Fluxions, 374, 388, 466
 focal property, 165–167
 focus, 165–167
 FOIL, xxv
 folding, 272
 folium of Descartes, 364
 foot, 30
 force, 476
 form, 188
 Platonic, 85
Formal Logic, 544, 546
 formalism, 550–551
 formula, xxv
 Guldin's, 195, 198
 quadratic, xxv, 147, 234
 Stirling's, 423
 well-formed, 555
 four-line locus, 165–166, 168, 193, 361
 Fourier coefficients, 514
 Fourier integral, 391, 512–513, 516
 Fourier inversion formula, 516
 Fourier series, 11, 391, 512–514, 519
 fourth proportional, 521
 fraction, 39, 59, 209, 242, 245–246
 continued, 101
 Horus-eye, 58, 62, 67
 improper, 245
 sexagesimal, 38
 fractional-linear transformation, 449
 France, 192, 318, 338, 376, 411, 423, 473
 Franks, 283
 French language, xxvi, 31, 298, 351, 352, 412, 452
 Frenet–Serret equations, 477
 frequency, 10
 frustum, 55, 74, 228, 252
 of a cone, 150
 of a pyramid
 volume, 49, 74
fu, 245
fukudai, 272
fukudai license, 272
 function, 8, 389, 391
 absolutely continuous, 528
 algebraic, 387, 457
 analytic, 389, 500–508, 511, 521
 continuous, 495
 definition, 512, 513, 536
 development, 548
 Dirichlet, 515
 doubly periodic, 504
 elementary, 382
 elliptic, 97, 386
 exponential, 347, 378, 500, 545
 generalized, 512
 harmonic, 425, 518, 528
 hyperbolic, 6, 484
 logarithmic, 347, 500
 “mechanical”, 495
 monotonic, 528
 multivalued, 457, 458, 505
 piecewise monotonic, 515
 quadratic, 306
 rational, 383, 438, 441, 500
 theta, 407, 504
 transcendental, 169, 457
 trigonometric, 6, 184, 221, 226, 308, 323, 378, 500
 wave, 427
 functional, 400
 functional analysis, 530, 537, 540
 fundamental form
 first, 471, 477, 479
 second, 477
 fundamental group, 460
 fundamental theorem of algebra, 433, 439–445, 461, 544
 fundamental theorem of calculus, 370
 furlong, 30

 gallon, 30
 Galois group, 445
 Galois theory, 4
 gambling, 9
 Ganesh, 209

- ganita*, 209
Ganitapada, 219
 GAR, 46
 Gauss–Bonnet theorem, 473
 Gaussian curvature, 470, 479
 Gaussian distribution, 427
Gebilde, 508
 geese, 233
 Gemini, 183
 general relativity, 394, 412
General Source of Computational Methods, 243
 generalized function, 512
 generalized hyperbola, 367
 Genesis, 294
 genus, 293, 503, 506
 geocentric astronomy, 181, 189
 geodesic, 394, 464, 472, 476
 geodesic curvature, 473
 geodesy, 469
Geography, 286
 geography, 78, 80, 177–189, 285
Geometria organica, 453
geometria situs, 456
Geometria prima elementa, 486
 geometric algebra, 105, 113, 142, 147
 geometric progression, 343
Géométrie, 360, 363, 521
 geometry, 5–6, 8, 13, 17, 25, 46–56, 66–77, 95, 98, 117, 128, 190, 201, 205, 210, 219–226, 255–266, 270–277, 302–310, 313, 323, 340, 386, 411, 532
 algebraic, 453–463, 481, 492
 analytic, 163, 166, 167, 199, 205, 288, 330, 337, 358–362, 373, 382, 448
 Chinese, 249–253
 descriptive, 448, 450
 differential, 448, 464–481, 486, 492
 Euclidean, xxv, 6, 68, 103–114, 172, 177, 201, 225, 252, 348, 367, 372, 376, 480
 Greek, 191, 199
 Hellenistic, 169–176
 hyperbolic, 307, 472, 478, 483, 487
 imaginary, 488
 metric, 88, 172
 metric-free, 160, 177, 193, 199, 201
 non-Euclidean, 13, 137, 172, 309, 311, 448, 456, 472, 481–494
 plane, 86, 142, 220, 227–228
 projective, 348–357, 448–463, 481, 492
 Roman, 169–176
 solid, 86, 129, 144, 227–228, 262
 spherical, 220
 synthetic, 451
 German language, xxvi, 72, 452, 538
 Germany, 376, 387, 410, 412
Gesetz der Kanten, 459
 Gettysburg Address, 31
 Gibbs random field, 390
 Gibraltar, 283
 GIMPS, 95
 Girton College, 408
 global positioning system, 178
 gnomon, 41, 219, 242
Gnomon of the Zhou, 241
 goats, 57, 294
 Golden Ratio, 113
 Golden Section, 349
 goods, 425
 Göttingen, 410, 469, 477, 485
 Göttingen Royal Society, 477
 Göttinger Gesellschaft der Wissenschaften, 413
gou, 249
gougu theorem, 249, 252, 253
 government, 36, 85
 grain, 253
 gram, 430
Grammelogia, 345
 grandparent, 7
 graph, 169, 320
 graph theory, 17
 gravitation, 477
 gravitational equation, 479
Great Books of the Western World, 160
 great circle, 87, 219, 235, 308, 317, 484
 Great Pyramid, 82
 greatest common divisor, 101, 231
 greatest common factor, 92, 97
 greatest common measure, 92
 Greece, 6, 19, 28, 32, 77, 148, 197, 211, 319
 Greek civilization, 49
 Greek language, xxvi, 28, 57, 79, 81, 88, 89, 144, 161, 196, 203, 287, 292, 313, 317, 323, 417
 Greek mathematics, 72, 216, 283
 Greek numeration, 95
 Greenwich, 178
 gross, 30
 group, 4, 443, 444
 abelian, 443
 fundamental, 460
 homology, 460
 locally compact abelian, 539
Grundgesetze der Arithmetik, 549

- Grundlagen der Geometrie*, 492
gu, 249
 Gujarati language, 203
 Guldin's formula, 195, 198
 Guldin's theorem, 262

Habilitation, 412
Hai Dao Suan Jing, 243, 250
 half-chord, 118, 221
 Halle, 414
 Halys River, 82
 Han Dynasty, 201, 239, 241, 242
Handbook of Political Fallacies, 543
 Hannover, 376, 469, 493
 Hanoi, 179
 Harappa, 203
 harmonic function, 425, 518, 528
 harmonic series, 331, 383
harpedonáptai, 68
 Harran, 287, 288
 harvest, 68
hau computation, 64
 heat equation, 390, 399, 519
 Hebrew language, 285, 286
 hectare, 30
 Heidelberg, 334, 406
 height, 205
hekat, 67, 74
 heliocentric astronomy, 182, 189
 Helios, 115
 helix, 326
 Hellenic Era, 79
 Hellenistic civilization, 57
 Hellenistic Era, 79, 87, 197, 482
 Hellenistic mathematics, 286
 hemisphere, 68, 80, 235
 area, 72
 northern, 223
 heptagon, 122, 286
 heptagonal number, 95
 heptakaidecagon, 122
 heritage, 3, 4, 11, 147
 Heron's formula, 172, 227
 hexagon, 48, 143, 191, 252, 349
 hexagonal number, 95
 hieratic, 57, 59, 95
 hieroglyphics, 57, 58, 65, 95
Higher Plane Curves, 473
 Hilbert basis theorem, 412
 Himalaya Mountains, 204
 Hindi language, 203
 Hindu mathematics, 101
 Hindu–Arabic numerals, 206, 292–293,
 328–329
 Hinduism, 203, 205, 208, 211, 213, 216, 285,
 287, 291, 292, 308
Hisab al-Jabr w'al-Muqabalah, 287
 historical ordering, 17
History of Herodotus, 68
 history, 486
 political, 3
 Hittite civilization, 28
 holes, 17
 Homeric poems, 204
 homogeneous coordinates, 454
 homologous curves, 459
 homology, 460
 homology group, 460
 homotopy, 557
 honeycomb, 191
 horizon, 224, 225
 horizontal, 180
 Horner's method, 258, 300
 horocycle, 490
Horologium oscillatorium, 464
 horosphere, 487, 490
 horse
 draft, 219
 thoroughbred, 219
 horseshoe, 544
 Horus-eye fraction, 58, 67
 Horus-eye parts, 62
 l'Hospital's rule, 379
 hour, 33
 House of Cancer, 183
 House of Wisdom, 286
 Huguenot, 423
 Hungary, 320
 hydrostatics, 150
 Hyksos, 56
Hypatia, or New Foes with an Old Face, 198
 hyperbola, 118, 126, 161–163, 165,
 171, 352
 branch, 164, 166
 equation, 164
 generalized, 367
 nappe, 119
 rectangular, 121, 297
 hyperbolic cosine, 484
 hyperbolic function, 6, 484
 hyperbolic geometry, 307, 472, 483, 485, 487
 hyperbolic paraboloid, 479
 hyperbolic plane, 478
 hyperbolic sine, 484

- hyperelliptic integral, 407
hypotenuse, 47, 53, 152, 173, 185, 198,
221, 257
- i*, 496
- Iceland, 178
- icosahedron, 144, 445
- idea, 188
Platonic, 85
- idempotence, 548
- ideogram, 39
- Iliad*, 107, 117
- Ilkhan*, 290
- Illinois, 410
- Illustrated London News*, 15
- imaginary geometry, 488
- imaginary number, 340, 495–496, 500
- incidence, 492
- inclusion-exclusion principle, 419
- incommensurable, 130, 139
- incommensurables, 86, 92, 106–113, 116, 146,
372, 384
quadratic, 141, 144
- incompleteness theorem, 542, 554–555
- independent trials, 419, 428
- independent variable, 514
- indeterminate problem, 99
- India, 6, 20, 22, 28, 30, 46, 201, 203–213, 218,
233, 239, 267, 274, 275, 283, 284, 290,
292, 310, 358
- Indian languages, 204
- Indian Statistical Institute, 210
- Indiana, 70
- indivisible, 363
- Indo-European language, 27, 28, 205
- induction, 546
transfinite, 535
- Indus River, 203, 204
- inequality
isoperimetric, 176, 186, 191
- infimum, 440
- infinite, 137, 476
- infinite number, 217
- infinite precision, 6, 337
- infinite series, 157, 205, 258, 277
- infinitely infinite space, 217
- infinitesimal, 206, 237, 271, 274, 312, 358, 376,
378, 397, 400, 467, 500
- infinity, 8, 206, 211, 216, 353, 367, 451
actual, 217, 369
potential, 217
- inflection point, 467
- inheritance, 287
- initial condition, 390
- inscribed circle, 278
- Institute for Advanced Study, 414
- Institutiones calculi*, 382
- insula*, 174
- insurance, 9, 425, 427
- integer, 6, 38, 59, 109, 217
- integer part, 34
- integers
sum of initial segment, 44
divisibility, 91, 100
- integral, 358, 374, 377, 382, 511–520
abelian, 407, 444, 503
algebraic, 500–504
elliptic, 407, 445, 502
Fourier, 391, 512–513, 516
hyperelliptic, 407
Lebesgue, 515, 531
nonelementary, 382
Riemann, 518, 531
- integrating factor, 388
- integration, 383, 397, 437, 500
complex, 504
Lebesgue, 541
- interest, 248
- intermediate-value property, 461
- International Congress of Mathematicians,
240, 413
- interpolation, 30
- Introductio in analysin infinitorum*, 382, 500
- Introduction to Mathematical Studies*, 269
- Introduction to Plane and Solid Loci*, 361
- Introduction to Set Theory*, 554
- intuition, 316, 511
- intuitionism, 9, 551–553
- invariance, 145
of dimension, 556
- invariant, 455
- involute, 464, 479
- involution, 209
- Ionia, 77, 80, 84, 115, 126
- Iran, 288
- Iraq, 27, 109, 178, 284, 285, 289, 292
- Ireland, 315, 379
- irrational number, 6, 12, 109–113, 216, 302,
329, 555
- irrational root, 235
- irrigation, 248, 284
- Isis and Osiris*, 71
- Islam, 204, 240, 266, 281, 283–311, 313,
318, 482

- isoperimetric inequality, 169, 176, 186, 191
 isoperimetric problem, 78, 191, 469, 476
 isosceles triangle, 278
 isotherms, 473, 476
Istituzioni analitiche, 382
 Italian language, 20, 478
 Italy, 77, 320, 328, 338, 349, 408, 477, 478
- Jabal Tarik, 283
 Jacobi inversion problem, 504, 507
jadhr, 295
 Jainism, 204, 206, 211, 216–218, 225
 Japan, 201, 240, 267–280, 310, 447
 Japanese names, 267
jayb, 305, 318
 Jena, 548
 Jeremiah (book of the Bible), 28
 Jerusalem, 28
 Jesuit, 195, 240, 243
 Jews, 197
Jinkō-ki, 269, 273
Jiu Zhang Suan Shu, 242–244, 247–253, 256, 262, 265
Jiu Zhang Suanshu, 201
jiva, 221, 305, 318
Journal de l'Ecole Polytechnique, 459
Journal für die reine und angewandte Mathematik, 406, 442
Journal of the Warburg and Courtauld Institute, 149
 Judah, 28
 Julian calendar, 406
 Jupiter, 178, 181
jya, 305
- Kō*, 267
Kaballah, 19
Kai Fukudai no Ho, 272
 Kaiser, 412
 Kaliningrad, 457
kalpa, 217
Kalpa Sutras, 217
kanji, 267
 Kansas State Agricultural College, 410
 karat, 30
kardo maximus, 174
 Kattigara, 179
Katyayana Sutra, 215, 216
 Kazan' Physico-Mathematical Society, 487
Ketsugi-shō, 275
khar, 73
khet, 69
- Khorasan, 288
 Kievan Rus, 308
 Kingdom of Wei, 242
 Kings (books of the Bible), 28, 70
Kitab al-Manazir, 305
Kitab al-Zij, 288, 318
 knitting, 17
 knot, 13
 knowledge
 - a posteriori*, 7
 - a priori*, 7, 541
 - analytic, 7
 - synthetic, 7, 541*Kokon Sampō-ki*, 271
 Königsberg, 320, 397, 457
 Königsberg bridge problem, 17, 457
 Korea, 239, 267, 268, 271
koti, 219
 Kuba, 17
kun reading, 267
 Kusumapura, 207, 208
kuttaka, xxiv, 224, 229–233, 238, 254
- L'invention nouvelle en l'algèbre*, 434
La perspective de Mr Desargues, 354
 Lagrange multiplier, 394
 Lambert quadrilateral, 303
 language, 18
 - Indo-European, 205
 - invented, 542
 Laplace's equation, 513, 519
 Laplace–Beltrami operator, 478
 Laplacian, 478, 513
 Larsa, 38
 last ratio, 397
 Latin alphabet, 218
 Latin language, xxvi, 79, 81, 144, 161, 178, 226, 285, 286, 294, 305, 313, 323, 352, 374, 377, 393, 456
 Latin square, 255
 latitude, 174, 177–180, 188, 224, 225, 235, 478
 latitude of forms, 337
latus rectum, xxv, 162, 163, 165, 168
 Laurent series, 505
 law, 5, 287, 296, 376
 - excluded middle, 9
 - Roman, 291
 law of cosines, 337
 law of large numbers, 419, 422, 428–429
 - weak, 429
 law of sines, 308, 332, 362
Laws, 86

- Laws of Thought*, 546–548
Le progrès de l'est, 408
 league, 30
 “leaning-ladder” problem, 47, 54
 least action, 476
 least common multiple, 97
 least-squares, 387, 426
 leather roll, 57
 Lebesgue integral, 515, 526, 531
Leçons sur le calcul des fonctions, 500
Leçons sur les fonctions discontinues, 529
Lectiones geometricae, 370
 leg, 152, 185, 198, 257
 legacy, 296–297, 300, 301
 legislated value of π , 70
 Leipzig, 320, 376, 377
 lemma, 145
 lemniscate, 501, 502
 length, xxv, 6, 19, 30, 41, 42, 46, 69, 129, 142,
 150, 172, 176, 184, 194, 195, 205, 224,
 242, 249, 526
 lens grinding, 275
Let's Make a Deal, 23
 lever, 150, 198
 Leyden, 208, 360
li, 247, 251, 253
Liber abaci, 316, 324–326
Liber calculatorum, 331
Liber de ludo, 419
Liber quadratorum, 326, 341
 liberal arts, 241
 liberal education, 209
 Library at Alexandria, 116, 140, 196, 286
 Libya, 197
 Liège, 316
 life expectancy, 418
 light, 392, 477
 light ray, 137, 166
 light-year, 488
 lightning, 20
Lilavati, 209, 233, 238
 limit, 153, 157, 375, 381
 limit point, 533
 line, 84, 85, 96, 106–108, 122, 136, 137, 160,
 162, 169, 176, 192, 199, 211, 220, 298,
 332, 350, 359
 directed, 487
 infinite, 353, 357
 transversal, 139
 line at infinity, 451, 484
 line coordinates, 455
 linear algebra, 247
 linear dependence, 474
 linear equation, 40, 230, 272, 296
 linear independence, 474
 linear number, 96
 linear system, 453
 lines
 converging, 307
 parallel, 136, 171, 224, 340, 352, 353,
 462, 484
 perpendicular, 177, 220
 linkage, 310, 361
 Lisieux, 288, 319
 literature, 217
Lives of Eminent Philosophers, 81, 82, 197
 Lo River, 255
 loan, 248
 local solar time, 178
Loci, 140
 locus, 124, 162, 191, 192, 199, 302, 359
 five-line, 193
 four-line, 161, 162, 165–166, 168, 193, 361
 plane, 359, 379
 six-line, 193, 359
 solid, 359
 three-line, 161, 165, 168, 193, 361
 two-line, 167, 193
locus, 456
 log, 279
 logarithm, 31, 169, 323, 338, 343–344, 346,
 378, 388
 Briggsian, 31, 344
 hyperbolic, 423
 natural, 425
 logarithmic function, 347, 500
 logic, 6, 7, 13, 21, 211, 216, 315, 360, 427,
 542–557
 three-valued, 553
 logical relation, 8
 logicism, 7, 549
 London, 225, 345, 376, 381, 408
 London Mathematical Society, 410
 longevity, 421
 longitude, 174, 177, 179, 180, 188, 235, 478
 lottery, 24
 lotus, 59
 Louvre, 39, 41
 Lower Egypt, 56
 Lower Saxony, 469
 lowest terms, 97
L_p-spaces, 527
 lune, 116, 138, 144
Luo-chu-shu, 255

- Luo-shu*, 255
 Luxor, Egypt, 57
 Lyceum, 116, 135
 Lydia, 82
 Lyons, 320

 Maclaurin series, 382, 388
 MacTutor website, 229
 Madagascar, 19
 Madras, 211
 magic square, 19, 255
 magnetism, 477
 magnitude, 302, 313
Mahabharata, 203, 233
 Mainardi–Codazzi equations, 477
 Mainz, 376
 major axis, 372
 major fifth, 13
 Malagasy, 19, 24
 al-Mamun, 284
man height, 205
 Manchu (Ching) Dynasty, 240, 244
 manifold, 464, 475, 534
 flat, 476
 map, 177, 179
Maple, 6
 mapping
 conformal, 464, 469, 473
 degree-preserving, 449
 Maragheh, 290
 Marburg, 485
 market, 425
 Markov chain, 429
 marriage, 71, 406, 415
 Mars, 181, 188
 Masillia (Marseille), 178
Mathematica, 6, 45, 263
Mathematical Analysis of Logic, 546
 Mathematical Association of America,
 415, 479
Mathematical Classic of Sun Zi, 242
 mathematical cranks, 126
 mathematical expectation, 424
 Mathematical Institute, 412
 mathematical logic, 8
 mathematical reasoning, 8, 20–22
mathēmatikoí, 84
Mathematische Annalen, 413, 478, 491
Mathematische Keilschrifttexte, 29
 matrix, 256, 258
 transition, 429
Matsya Purana, 208

 maximum, 363
 Maya, 197
 mean and extreme ratio, 142, 349
 mean position, 183
 mean proportional, 97, 118, 123, 128, 143, 147,
 166, 497
 mean solar time, 31
 means, 143
 measure, 242, 535
 Borel, 539
 measure of curvature, 470
 measure of precision, 427
 measure space, 422
 measure theory, 411, 537
 measure zero, 528
 measurement, 6, 68, 80, 172, 320
Measurement of a Circle, 150, 172
 Mecca, 285, 290
 mechanical drawing, 451
 mechanics, 89, 150, 288, 314, 319, 359, 383,
 386, 389, 476
 Medes, 82
 medicine, 217, 320, 411
 Medieval Era, 311–323
 Mediterranean Sea, 27, 28, 79, 201, 203,
 283, 318
 membrane, 511
*Memoir on Some Traditions of the
 Mathematical Art*, 255
*Memorandum for Friends Explaining the Proof
 of Amicability*, 294
 Menelaus' theorem, 353, 356
Mengenlehre, 530
Meno, 47, 111
 Mercury, 181
 meridian, 326
 prime, 178
 meridian of longitude, 179
 Merôe, 179
 Mersenne prime, 95
 Merton College, 320, 331
 Merton rule, 320, 331
Meru Prastara, 218
 Mesopotamia, 6, 19, 22, 25, 27–55, 65, 76, 79,
 82, 84, 90, 98, 138, 142, 203, 211, 216,
 274, 283, 290, 292
 metalanguage, 551, 554
 metamathematics, 9
Metaphysics, 109
 metaphysics, 8, 85, 211, 216, 239
 meter, 234
Method, 150, 153–155, 263, 266, 365

- method of exhaustion, 131–132, 153, 217, 235, 365, 367–368, 376, 397
 method of indivisibles, 363, 384
Method of Interpolation, 243
Method of Solving Fukudai Problems, 272
Methodus inveniendi lineas curvas, 393
 metric
 Cayley, 456
 p -adic, 530
 projective, 456
 Riemannian, 478
 metric geometry, 78
 metric space, 530
 metric system, 30, 32
 metric-free, 235
Metrica, 154, 172
 microgram, 430
 Middle Ages, 96, 148
 Middle East, 30, 79, 81, 187
 Middle Kingdom, 56
 midpoint, 125, 139, 309
 Miletus, 80
 millet, 247
 Ming Dynasty, 204, 240
 minimal surface, 394, 469
 minimum, 363, 440
 minus of minus, 496
 minute, 30, 33, 236
Mirifici logarithmorum canonis descriptio, 323, 343
Miscellanies, 68
 missionary, 243
 Möbius band, 452, 459
 Möbius transformation, 449, 452
 model, 482
 modern algebra, 4
 modulus, 110
modus ponens, 548
 Mogul Empire, 204
 Mohenjo Daro, 203
 monads, 84
 monasteries, 313, 315
 monastery schools, 313
 Mongol Empire, 240, 284
 Mongols, 240, 284, 289
 month, 12, 30, 230, 248
 moon, 48, 180
 full, 230
 phases, 178
 moral certainty, 422
 Moravia, 15
 Morley's theorem, 199
 Morocco, 216
 Moscow, 74, 406, 477, 535, 540
 Moscow Museum of Fine Arts, 57
 Moscow papyrus, 72, 74, 252
 mosque, 285
 motion, 162, 169, 196, 304, 307, 313
 retrograde, 188
 uniformly accelerated, 320, 331
 Mount Meru, 218
 Mount Olympus, 218
 Mozambique, 179
 multilinear algebra, 474
 multilinearity, 272
 multiplication, 12, 34, 38, 59, 199, 209, 245, 292, 322
 Egyptian, 61
 multiplication table, 38
 multivalued function, 458, 505
 Mumbai (Bombay), 210
 Murchiston, 323, 343
 Museum of Alexandria, 196
 music, 9–11, 210, 242, 313, 315
Musica Humana, 315
Musica Instrumentalis, 315
Musica Mundana, 315
 Muslim civilization, 57, 207
 Muslim conquest, 81
 mustard seed, 211
 mutual-subtraction algorithm, 264
 mythology
 Greek, 218
 Hindu, 218

 n -gon, 122
 Nagoya, 277
naka, 267
 nameable number, 211
 nappe, 119, 164
Natural History, 178
 natural logarithm, 425
 natural number, 84
 Navarre, 341
 Navier–Stokes equations, 473
 navigation, 80
 Nazism, 413
 necessity, 427
 negative number, 216, 234, 245, 285, 321, 336, 340, 500
 square root of, 322
neûsis, 124, 127, 297
 New Kingdom, 56
 New Testament, 81

- New York Historical Society, 57
 Newnham College, 409
 Newton's laws, 464
 Newton's method, 384
 Newton–Raphson approximation, 36, 39
Nicomachean Ethics, 293
 Nile River, 57, 116, 179
Nine Chapters on the Mathematical Art, 242
Nine Symposium Books, 103
Nine-Chapter Mathematical Treatise, 201
 Noetherian ring, 413
 non-Euclidean geometry, 13, 137, 311, 472, 481–494
 nonorientable surface, 459
 nonstandard analysis, 376, 399
 normal distribution, 427, 428, 430
 normal form, 391
 normal subgroup, 444
 North Africa, 81
North China Herald, 247
 North Pole, 179
 northern hemisphere, 223
 Norway, 382
 notation, 338, 340
 notebooks, Ramanujan's, 211
 nothing, 547
 number, 6, 8, 25, 172, 313
 algebraic, 295, 523, 534
 Avogadro, 430
 Bernoulli, 422
 cardinal, 534
 complex, 340, 346, 379, 389, 445, 448, 453, 495–496, 500, 513
 composite, 93
 critical, 552
 deficient, 94
 enumerable, 211, 217
 equilateral, 110
 even, 72, 93, 255
 evenly even, 93
 figurate, 77, 91, 95–118
 floating-point, 6
 heptagonal, 95
 hexagonal, 95
 imaginary, 340, 495–496, 500
 interpretation, 498
 infinite, 206, 217
 irrational, 6, 12, 109–113, 302, 329, 555
 linear, 96
 nameable, 211
 natural, 84
 negative, 216, 234, 245, 285, 321, 336, 340, 500
 oblong, 110
 odd, 72, 93, 113, 255
 ordinal, 532, 533, 540, 549
 countable, 551
 pentagonal, 91, 95, 109
 perfect, 72, 91, 94, 97, 102, 293
 plane, 96
 polygonal, 97
 polyhedral, 95
 positive, 295
 prime, 93, 122, 553
 rational, 98, 109, 216, 245, 295, 327, 555
 real, 6, 109, 193, 199, 288, 295, 337, 340, 347, 383, 445, 499, 511–523
 square, 72, 91, 95, 98
 superabundant, 94
 transcendental, 445, 522, 536
 triangular, 91, 95, 109, 316
 unenumerable, 217, 225
 number system, 32
 Chinese, 32, 244–246
 sexagesimal, 33–35
 conversion, 33–35
 number theory, 6, 8, 31, 77, 100, 143, 190, 228–233, 238, 255, 288, 292–294, 386
 Greek, 91–102
 numbers
 amicable, 293
 divisibility, 97
 relatively prime, 93, 97
 numerals
 Hindu–Arabic, 204, 206, 284, 292–293, 316, 319, 328–329
 numeration system, 213
 numerology, 24
 Nürnberg, 320, 412

 oblong number, 110
 observational error, 426
 obtuse angle, 308
 octahedron, 144
 octave, 13
 odd number, 72, 93, 110, 113, 255
Odyssey, 94
Oedipus the King, 117
oikuménē, 179
 Old Babylonian language, 27, 38, 46, 48, 55, 58, 65
 Old Kingdom, 56
 Old Persian language, 29

- On Burning Mirrors*, 169
On Exile, 117
On General Triangles, 331
On Isoperimetric Figures, 169
ON reading, 267
On Socrates' Daemon, 130
On the Pythagorean Life, 84
 one-dimensional π , 216, 219, 227, 251, 262, 264
 one-dimensional space, 217
 one-sided surface, 459
 operation
 algebraic, 5
 Opium War, 240
 opposition, 189
 optative mood, 417
Optics, 87, 140, 171
 optics, 150, 286, 359, 394, 406
 oracle, 19
 order, 353, 355
 ordering, 535
 ordinal number, 532, 533, 540, 549
 countable, 551
 ordinate, 163, 164, 367, 377, 389
 organic chemistry, 350
Origine, trasporto in Italia, primi professi in essa dell' algebra, 295
 ornamental geometry, 285
 orthogonal vectors, 474
orthotomē, 119
 ostracon, 80, 141
 Ottoman Empire, 284, 376
 outer product, 474
 Owari, 277
 oxen, 57
 Oxford, 294, 320, 331
 Oxyrhynchus, 141
oxytomē, 119

 pagans, 197
 Paingloss, Dr., 375
 painting, 323, 348–352, 357, 448
 Pakistan, 201, 203, 206, 209
 Palermo, 326, 478
 Palestine, 28, 81, 284
 palimpsest, 149
 palmistry, 19, 24
 panda, 7
 Pappus' theorem, 81, 193
 papyrus, 57, 66, 80, 141, 244
 Moscow, 72, 74
 Reisner, 57
 Rhind, 22, 57–65, 69
 parabola, 118, 126, 161–163, 165, 352, 363–364, 367, 479
 quadrature, 150, 155–157
 segment, 87
 paradox, 139, 461, 511
 arrow, 375
 Banach–Tarski, 540, 541
 Burali-Forti, 539, 549
 Cramér's, 463
 Petersburg, 424–425
 Russell's, 537, 549, 550, 553
 Zeno's, 107–108, 113
 Achilles, 107
 arrow, 108
 dichotomy, 107
 stadium, 108
 parallax, 488, 493
 parallel lines, 136, 165, 224, 352, 353, 462, 484
Parallel Lives, 81
 parallel of latitude, 177
 parallel postulate, 13, 103, 171–172, 193, 286, 302–309, 448, 481
 parallelepiped, 144, 194
 parallelism, 492
 parallelogram, 104, 113, 134, 141, 191, 198
 infinitesimal, 471
 parameter, 426, 477
 parameterized surface, 468
 Paris, 39, 41, 317, 320, 330, 354, 376, 386, 408, 420
 Parma, 295
Parmenides, 134
 parrots
 counting ability, 15
 parsec, 488
 part (divisor), 94
 part (unit fraction), 76
 double, 65, 76
 partial derivative, 520
 partial differential equation, 390–391, 425
 normal form, 391
 partial fraction, 437, 438
 partial reinforcement, 20
 parts (unit fractions), 58, 59, 62–65
 double, 62
 Pascal's theorem, 355, 449, 454, 462
 Pascal's triangle, 218, 434
 Pasch's theorem, 303
 Pataliputra, 207
 Patna, 207

- Paul (Alexandrian astrologer), 206
Paulisha Siddhanta, 206
 Peaucellier linkage, 124
 pebbles, 18, 38
 peck, 30
 pedagogical ordering, 17
 pedagogy, 54, 243, 451
 Peleset, 28
 Pell's equation, 229, 237
 greater solution, 230
 lesser solution, 230
 Pella, 29
 Peloponnesian War, 28, 115, 116
 Peloponnesus, 115
 pencil, 353
 pendulum, 383, 407, 501
Pensées, 355
 pentagon, 113, 122, 130, 143, 144, 146,
 185, 349
 pentagonal number, 91, 95, 109
 pentakaidecagon, 143
 people, 246
 perfect number, 91, 94, 97, 102, 293
 perfect set, 535
 Perga, 161
 Pergamon Museum, 43
 Pergamum, 88
 perigee, 182
 perihelion, 182
 perimeter, 19, 48, 169
 Period of Warring States, 239
 permutation, 5, 233, 421, 439, 442, 444
 perpendicular lines, 177
 Persia, 283, 284, 289, 290
 Persian Empire, 28, 115, 116, 178
 Persian Gulf, 27, 28
 Persian language, 204, 285, 288
 perspective, 348, 350, 352, 450
 Peshawar, 206
pesu, 67, 69, 76, 247, 248
 Petersburg paradox, 424–425
Phænomena, 87, 140, 286
 pharaoh, 18
 pharmacy, 335
 phase, 10
 Φ , 12, 113, 325, 337
 Philistines, 28
 philosophy, 3, 6, 77, 80, 129, 198, 241, 286,
 291, 314, 359, 384, 400, 405, 427–428
 Greek, 82
 neo-Platonic, 86, 91, 190, 198
 Platonic, 85–87, 97
 pre-Socratic, 81
 Pythagorean, 91, 97, 115
 Stoic, 171
 phyllotaxis, 325
Physics, 107, 131
 physics, 80, 196, 307, 412, 425, 476–477, 479,
 495, 510, 511
physikoi, 84
 π , 6, 12, 30, 48, 55, 227, 251, 276, 279, 318,
 370, 423, 484, 522
 “biblical” value, 70
 Archimedes' estimate, 153
 Egyptian value, 70
 irrationality, 484
 legislated value, 70
 “neat” value, 227
 one-dimensional, 48, 73, 216, 219, 227, 251,
 262, 264
 three-dimensional, 219, 265
 transcendence, 445, 522
 two-dimensional, 73, 216, 219, 227, 251
piece, 298, 301
 pigeons, 20
 pint, 30
 pipe, 248
 Piraeus, 130
 Pisa, 319
 pitch, 315
 \pitchfork (*pitchfork*), 98
 place-value system, 32–33, 37, 38, 59, 65, 109,
 207, 212, 213, 292
 plague, 117
 planar problem, 123, 193
 Planck's constant, 6
 plane, 84, 96, 107, 144, 169, 191, 348
 complex, 449, 452
 projective, 459
 plane geometry, 86
 plane locus, 359, 379
 plane number, 96
 plane region, 195
 plane trigonometry, 308
 planet, 48, 180, 181, 208, 418
 planting, 68
Platonicus, 117
 Platonism, 85
 Playfair's axiom, 171
 Plimpton 322, 49–54
 Plimpton collection, 49
 plucked string, 512
 plus of minus, 496
 Poincaré conjecture, 461

- point, 96, 107, 108, 162
 - cluster, 533
 - limit, 533
- point coordinates, 455
- point of accumulation, 533
- point-set topology, 534
- points
 - collinear, 354
 - conjugate, 395
- pointwise convergence, 525–526
- Poisson distribution, 428
- Poland, 406
- polar coordinates, 368, 518
- polarization identity, 48, 142
- pole, 507
- polis*, 135
- political history, 3
- politics, 115
- polygon, 68, 84, 104, 116, 131, 141, 145, 150, 216
 - 17-sided, 443
 - circumscribed, 143
 - inscribed, 143
 - of 192 sides, 252
 - regular, 169, 220
- polygonal number, 97
- polygons
 - similar, 143, 198
- polyhedral number, 95
- polyhedron, 84, 262, 458–459
 - closed, 459
 - regular, 150
- polynomial, 98, 169, 258, 383, 445, 453
 - cubic, 407
 - irreducible, 445
 - prime, 445
 - quadratic, 407
 - symmetric, 434
- polytheism, 205
- pope, 318
- population, 253
- Porisms*, 140
- Portugal, 244
- positive number, 295
- postulate, 244
 - parallel, 171–172, 193
- pound, 30
- Power Ball, 432
- power series, 210, 382, 398, 495, 526
- power set, 537
- Practica geometriae*, 317, 323
- Prague Scientific Society, 442
- Prasum, 179
- prayer, 285
- pre-Socratic philosophy, 81
- precision
 - infinite, 6
- predicate, 135
- predicate calculus, 549
- Pregel River, 457
- premise, 9
- prime, 51
 - Mersenne, 95
- prime decomposition, 294
- prime meridian, 178
- prime number, 93, 122, 553
- prime numbers
 - infinitude, 97
- Princeton University, 414
- Principia mathematica* (Newton), 373, 374, 393, 467
- Principia mathematica* (Whitehead–Russell), 550
- Principia mathematica*, 8
- Principles of Mathematics*, 550
- Prior Analytics*, 135
- prism, 74, 131, 144, 191
- prisoners, 57
- Privatdozent*, 407, 412
- probability, 8, 9, 311, 379, 386, 417–432, 526, 546
 - conditional, 23, 419
- probability space, 422
- problem
 - “leaning-ladder”, 47, 54
 - planar, 193
 - Sturm–Liouville, 513, 515
 - vibrating string, 519
- product, 38, 40, 142, 361
 - infinite, 370
- projection, 224, 456
 - conical, 180
- projective geometry, 348–357, 492
- projective metric, 456
- projective plane, 459, 480
- proof, 13, 82, 327
- proof by contradiction, 552
- proper class, 553, 555
- proportion, xxv, 12, 66, 68, 82, 97, 101, 109, 112, 130–134, 141, 142, 144, 147, 150, 160, 163, 176, 177, 194, 195, 242, 248, 254, 397
- Eudoxan theory, 143

- proposition, 244
 universal, 85
 propositional calculus, 544
prosthaphæresis, xxiv, 4, 333–335, 337, 346
 Provence, 319
 pseudosphere, 478, 486, 492
Psychologie als Wissenschaft, 451
 psychology, 3, 14, 16, 539
 Ptolemais, 197
 Ptolemy's theorem, 184
 Pulkovo Observatory, 488
 pulley, 149
 pulverizer, 224, 229, 231
 purana, 208
 pure magnification, 469
 pure mathematics, 8
 Putnam Examination, 415
 pyramid, 6, 55, 68, 74, 80, 83, 131, 144, 228,
 252, 265
 Pythagorean comma, 13
 "Pythagorean" geometry, 103–106
 Pythagorean theorem, 6, 46–48, 67, 71–72, 81,
 84, 104, 113, 141, 142, 152, 185, 191,
 198, 201, 214, 243, 249, 252, 271, 309,
 487, 492
 generalizations, 143, 304
 Pythagorean triple, 213, 229
 Pythagoreanism, 315
 Pythagoreans, 19, 81, 103, 113, 128, 197

qian, 248
 Qin Dynasty, 239
 quadrant, 235
 quadratic equation, 19, 40, 69, 99, 105, 228,
 233, 256, 272, 296, 301, 329, 338, 359,
 433, 447
 quadratic form, 456
 quadratic formula, xxv, 147, 234, 260
 quadratic function, 306
 quadratic incommensurables, 141, 144
 quadratrix, 107, 117, 123, 124, 126, 169, 361,
 364, 377, 453
 quadrature, 367
Quadrature of the Parabola, 149
 quadric surface, 450, 454
 quadrilateral, 184, 227, 273
 area, 227
 cyclic, 173, 227
 Lambert, 303
 Saccheri, 303, 309, 482
 semi-Thabit, 303
 Thabit, 303, 307–309, 482

 quadrilateral problem, 274, 279
 quadrivium, 210, 313, 315
 quaestor, 148
 quantic, 456
 quantity
 continuous, 5, 6
 quantum mechanics, 427
 quartic, xxv
 quartic equation, 234, 260, 261, 308, 322,
 338–340, 346, 433, 438
 quartic polynomial, 407
 quaternion, 447, 479
Quatrains, 289
 quinquenove, 421
 quintic equation, 4, 439, 442, 445
 quotient, 34, 92, 101, 231, 329

 radial coordinate, 181
 radian, 221
 radian measure, 30
 radiator, 76
 radical, 234
 radio, 178
 radioactive decay, 432
 radium-228, 430
 radius, 48, 49, 79, 117, 123, 144, 151, 179,
 184, 185, 216, 227, 236, 261, 275, 278,
 298, 372
 imaginary, 484
 radius of curvature, 466, 479, 483, 487
 principal, 470
radix, 335
 railroad, 309
 rainbow, 286
Ramayana, 203
 random variable, 418, 428
 al-Raqqa, 288
 ratio, xxv, 12, 92, 97, 109, 112, 130, 133, 145,
 147, 176, 195, 199, 216, 220, 288, 332,
 354, 362
 anharmonic, 351
 composite, 133, 138, 145, 193, 194
 duplicate, 133
 final, 374
 initial, 374
 mean and extreme, 142
 rational function, 383, 438, 441, 500
 rational number, 98, 109, 216, 245, 295,
 327, 555
 ravens
 counting ability, 15
 real analysis, 387, 511–531

- Real and Complex Analysis*, 495
 real number, 6, 109, 193, 199, 288, 295, 337,
 340, 347, 383, 445, 499, 511–523
 real numbers
 completeness, 525–527
 real variable, 500
Recherches sur la probabilité des jugemens, 427
 reciprocal, 38
 terminating, 38, 50
 reciprocals, 38
 recombination, 264
 rectangle, 6, 19, 21, 41, 49, 68–69, 74, 104,
 138, 142, 152, 163–166, 179, 184, 194,
 214, 247, 249, 251, 303
 infinitesimal, 377
 rectangular coordinates, 368, 518
 rectangular hyperbola, 121, 297
 rectification, 369
 recursion, 225
 reflection, 511
 refraction, 306, 393, 511
 regular solid, 144, 349
 Reims, 316
 reincarnation, 208
 Reisner papyrus, 57
 relation
 logical, 8
 relative minimum, 299
 relative rate, 373
 relatively prime, 51, 93, 97, 113
 relativity, 476
 religious rites, 30
 remainder, 34, 92, 101, 254
 Renaissance, 349, 357, 448
Republic, 71, 86, 128
 residue, 505
 resolvent, 339, 438, 439
 retrograde motion, 161, 188
 revolution, 169, 193
Revue scientifique, 408
 Rhind papyrus, 22, 57–76, 242, 247, 251
 Rhodes, 179
 rhythm, 10
 rice, 247
 Riemann integral, 518, 531
 Riemann mapping theorem, 395
 Riemann surface, 458, 461, 506
 Riemann–Roch theorem, 506
 Riemannian manifold, 478
 Riesz–Fischer theorem, 412
 right angle, 122, 146, 185, 220, 303, 308
 right ascension, 180, 183
 right triangle, 152, 216, 221, 247, 249, 257,
 261, 371
 rigid body, 307, 407
rigle des premiers, 336
 ring, 4, 443
 Boolean, 548
 Roman Empire, 29, 79, 87, 177, 302, 311, 323
 Eastern, 291
 Roman law, 291
 Roman numerals, 31
 Rome, 19, 32, 70, 148, 177, 313, 316, 320
 root, 5, 218, 240, 245–246, 265, 295, 322, 330,
 335, 433, 438
 rope fixers (surveyors), 71
 rotation, 302, 501
*Rough Draft of an Essay on the Consequences
 of Intersecting a Cone with a Plane*, 352
 Royal Society, 371, 376, 381
 RSA codes, 102
Rubaiyat, 289
 rug, 288
 rule of inference, 9
 Rule of Three, 19, 65, 67, 248, 324
 ruler, 444
 Russell's paradox, 537, 549, 550, 553
 Russia, 19, 240, 284, 289, 329, 406, 407,
 451, 457
 Russian language, xxvi
Rv, 335

 Sabian, 287, 288
 Saccheri quadrilateral, 303, 309, 482
 Sacramento, 466
sagitta, 275
 St. Gerald Monastery, 316
 St. Petersburg, 386, 406, 461, 513
 St. Victor Abbey, 317
 Sakhalin, 31
 Salamis, 115
 Samarkand, 287, 302
 Samos, 80, 84, 189
 SAN, 268
sanbob, 268
Sand-reckoner, 150, 211
sangaku, 202, 268, 274, 277–280
 Sanskrit language, xxvi, 203–205, 207, 217,
 218, 221, 284, 294, 305, 318
 Saturn, 181
 scaling, 184
 schema, 550
Science, 431
 science, 203, 211

- score, 31
 Scotland, 178, 322, 338, 343, 409
 scroll, 352
 sculpture, 323
 Scythians, 178
Sea Island Mathematical Manual, 242, 250
Sea Mirror of Circle Measurements, 271
 Sea Peoples, 28
 secant, 143, 221, 226, 305
 second, 30, 33, 186
 second category, 529
 second fundamental form, 477
 Second Punic War, 87, 148
 Section (Golden Ratio), 142, 143
 section (square mile), 30
 sector, 179
 segment, 367
 of a parabola, 87
 Segovia, 318
seked, 69, 76
 Selasca, 478
 Seleucid Kingdom, 29
 Seljuks, 284, 289
 seller, 425
 semantics, 544, 550
 semiaxis, 372
 semicircle, 82, 116, 185, 224, 262, 298, 330
 semidifference, 39–41, 44, 50, 51, 55, 142, 252
 semiregular solid, 349
 Semitic language, 27
 senator, 21
 Senkereh, 38
 senses, 217
 Sēres, 179
 series, 358, 372–374, 382, 383, 388, 464, 495,
 511–520
 binomial, 517
 Dirichlet, 525
 Fourier, 11, 391, 512–514, 519
 geometric, 358, 367, 530
 harmonic, 383
 infinite, 370
 Laurent, 505
 Maclaurin, 382, 429, 475
 power, 382, 398, 509, 526
 Taylor, 378, 379, 511, 517, 545
 trigonometric, 390, 509, 518
 uniqueness, 532–533
 Serpent, 418
 services, 425
 set, 162
 analytic, 536
 countable, 538
 derived, 533, 535
 perfect, 535
 uncountable, 535, 538
 set theory, 7, 8, 411, 448, 521, 528, 532–541
 descriptive, 535
 fuzzy, 547
 Seven Years War, 204
 sexagesimal notation, xxv
 sexagesimal number, 109
 sexagesimal system, 31, 37, 49, 83, 184,
 186, 292
 shadow, 82, 219, 250, 253
 Shang Dynasty, 239, 241, 244
 Shang numerals, 245
 Shanghai, 179
 shape, 6, 14, 16–18
 sheaf, 353
 sheep, 294
 Sheikh Abd el-Qurna, 68
 Shetland Islands, 178
 Shimura–Taniyama conjecture, 100
 Shintō, 202, 269, 277
 shoes, 540
 shogun, 270, 277
 shoot, 352
Shushu Jiyi, 255
 Sicily, 29, 77, 115, 128, 148, 178, 210, 283,
 286, 319, 324
siddhanta, 205, 206, 285
Siddhanta Siromani, 209, 235
 sieve of Eratosthenes, 93
 signal
 band-limited, 10
 silk, 247
 silkworm, 247
 silo, 70
 similar polygons, 132, 143
 similar triangles, 82, 261, 317
 simply connected, 458, 479, 505
 Simpson's paradox, 431
 simultaneity, 307
 Sind, 209
sind-hind, 285
 sine, 169, 188, 221, 225, 235, 236, 288, 305,
 323, 335, 343, 358, 393, 497, 523
 hyperbolic, 484
 sine wave, 10
 singularity, 506
sinistra decumani, 174
sinus, 305, 318, 323
situs, 456

- six-line locus, 193, 359
 size, 6
 skepticism, 360
 slide rule, 344, 345
 circular, 345
 slope, 67, 69–70, 363
 of pyramids, 69
 Smolensk, 414
 Snell's law, 306
 soccer, 350
 sociology, 3, 9
 socks, 540
 solar system, 483
 solid, 107, 195, 302
 of revolution, 154, 193, 262
 regular, 144
 revolution, 81
 solid geometry, 86, 129
 solid locus, 359
 solid problem, 123
 solution by radicals, 439–445
 Song Dynasty, 239, 255
 Sophist, 116
soroban, 268
 Soviet Union, 414
 space, 217
 measure, 422
 metric, 530
 three-dimensional, 459
 vector, 4
 space-filling curve, xxiii
 Spain, 81, 244, 283, 285, 286, 290, 318
 Sparta, 115
 Spartans, 28
 special relativity, 307
 specific gravity, 149
 speed, 6
 average, 196
 sphere, xxv, 6, 48, 68, 118, 144, 148, 150,
 169, 172, 208, 219, 262, 271, 274, 317,
 458, 486
 area, 150–155, 235–237, 279
 celestial, 177, 180, 224, 317, 479
 sector, 235
 segment, 251
 surface, 87
 volume, 150–155, 262
Sphere and Cylinder, 150
 spherical coordinates, 181
 spherical mapping, 470
 spherical triangle, 308, 317, 332, 484
 spherical trigonometry, 184, 305
 spheroid, 150
 spiral, 107, 150, 169, 360, 364, 374, 453
Spirals, 150
 spring, 223
 Springer-Verlag, 29
 square, 6, 17, 38, 40, 41, 46, 47, 50, 53, 71, 72,
 91, 95, 97, 98, 104, 109, 116, 118, 122,
 130, 132, 141, 146, 163, 164, 166, 176,
 191, 198, 213, 214, 216, 219, 227, 251,
 263, 279, 327, 384
 completing, 235
 doubling, 47, 214
 Latin, 255
 magic, 19, 255
 square root, 36, 38, 39, 41, 51, 67, 110, 144,
 185, 206, 216, 218, 219, 226, 245, 261,
 264, 336, 340, 346, 361
 irrational, 44
 second, 218
 third, 218
 squaring the circle, 106, 115–117, 131, 139,
 147, 153, 215, 275, 323, 377, 445
 stade, 179
 Stadium paradox, 108
 standard deviation, 423, 427
 standard time, 31
 standard unit, 6
 star, 48, 177, 180
 statics, 319
 statistics, 8, 9, 20, 417, 421, 427
 Steiner–Lehmus theorem, 199
 Steinmetz solid, 263
 Steklov Institute, 461
 stem, 352
 Step Pyramid, 56
Stetigkeit und irrationale Zahlen, 523
Sthananga Sutra, 206
 Stirling's formula, 423
 Stockholm, 407
 stonemasonry, 220
 story problems, 8
 straightedge, 186, 191, 484, 522
 streamlining, 393
 string, 279, 315
 plucked, 512
 vibrating, 400, 519
 string property, 166, 168
 strip, 279
Stromata, 68
 strong extremal, 397
 Sturm–Liouville problem, 513, 515, 519
Suan Fa Tong Zong, 243, 244, 269

- Suan Jing Shishu*, 241
suan pan, 268
Suan Shu Chimeng, 269
 subgroup, 444
 normal, 444
 subject, 135
 subjunctive mood, 417
 subtangent, 363, 370, 382, 389
 Fermat's construction, 363–364
 subtraction, 12, 31, 38, 59, 245
 successor, 549
Suda, 190, 196–198
Sulva Sutras, 205, 211, 213–216, 225
 sum of sines, 368
 Sumerian civilization, 27
 Sumerian language, 29, 39
*Summa de arithmetica, geometrica, proportioni
 et proportionalita*, 320, 335
 summer, 223
 summit angle, 482
 sun, 48, 180, 187, 189, 190, 223–225, 250, 251,
 253, 317
 orbit, 182–183
Sun Zi Suan Jing, 242, 245–247, 253
 sundial, 68, 249
 sunrise, 224
sunya, 207
 superabundant number, 94
 supernova, 418
 superposition, 304
 superstition, 21
 surface, 17, 96, 468–469
 curvature, 468
 curved, 80, 243
 flat, 468
 minimal, 464, 469
 nonruled, 73
 one-sided, 459
 parameterized, 468
 quadric, 450, 454
 Riemann, 461
 simply connected, 458
 surveying, 5, 30, 68–69, 174–175, 220, 221,
 241, 250, 285, 287, 317
 surveyor, 36
Surya Siddhanta, 206, 221
 Susa, 48
Sushu Jiu Zhang, 258
 Svayambhu, 208
 Switzerland, 411
 Syene, 179
 syllable, 234
 syllogism, 135, 360
 symbol, 18–20, 77, 91, 98–99, 218
 symbolism, 98
 symmetric polynomial, 436
 symmetry, 191
Symposium Discourses, 128
Synagogē, 89, 123, 140, 190–196, 359
 syntax, 544, 550
Syntaxis, 88, 89, 286
 synthesis, 192
 Syracuse, 87, 128, 148–159, 210
 Syria, 288, 289, 319
 Syriac language, 287
System of the Sun, 206

 table of chords, 184–187
 tablespoon, 30
 Taiwan, 240
taka, 267
 Talmud, 70
 Tang Dynasty, 239
 tangent, 53, 143, 165, 221, 226, 261, 305, 363,
 374, 376, 384, 455, 497
 Tarentum, 98, 128
 tarot cards, 19, 24
 Tata Institute, 210
tatamu, 272
 tautochrone, 393
 taxation, 253
 taxes, 248
 Taylor series, 379, 389, 511, 517
 Taylor's theorem, 398, 505, 545
 teaspoon, 30
 temperature, 391, 399, 511, 519
Ten Canonical Mathematical Classics, 241
tengen jutsu, 271
 tensor analysis, 478
 tensor product, 474
 tensor, curvature, 479
 terminating reciprocal, 38, 50
 tetrahedron, 74, 96, 106, 144, 219
 Thabit quadrilateral, 303, 307–309, 482
The Analyst, 379, 385
The Decline of the West, 414
The New Yorker, 10
*The Origins of Algebra and its Transmission to
 Italy and Early Advancement There*, 295
The Utility of Mathematics, 117
Theatetus, 110
 Thebes, 68
 Thebes, Egypt, 56
 theology, 5, 313

- theorem, 9, 13, 82
 Baire category, 530, 531, 540
 binomial, 373, 378, 421, 506
 Bolzano–Weierstrass, 533
 Brouwer fixed-point, 556
 Cantor–Bendixson, 535
 central limit, 428–429
 Desargues', 192, 354
 dominated convergence, 530
 Fermat's last, 89, 91, 100
 Fubini's, 541
 Gauss–Bonnet, 473
gougu, 249
 incompleteness, 554–555
 Menelaus', 353, 356
 Morley's, 199
 Pappus', 81, 193
 Pascal's, 355
 Pasch's, 303
 Ptolemy's, 184
 Pythagorean, 6, 46–48, 71–72, 81,
 104, 113, 141, 142, 152, 185, 191,
 198, 201, 214, 249, 252, 271,
 309, 492
 Riemann–Roch, 506
 Riesz–Fischer, 412
 Steiner–Lehmus, 199
 uniqueness, 145
 Whittaker–Shannon, 11
theorema egregium, 472
Théorie analytique de chaleur, 514
Théorie analytique des probabilités, 425
Théorie des fonctions analytiques, 500
*Theory of Functions of a Real
 Variable*, 526
 thermodynamics, 428
 theta function, 407, 504
thing, 20
 Thoulē, 178, 180
Thousand and One Nights, 284
 thread, 247
 three-body problem, 407
 three-dimensional π , 219, 265
 three-dimensional space, 217
 three-line locus, 168, 193, 361
 three-valued logic, 553
 thumb, 59
 thunder, 20
tian yuan shu, 271
 Tigris River, 27, 284
 tiling, 191
Timaeus, 96
 time, 6, 12, 30, 31, 208, 224, 320, 374,
 389–390, 401
 local solar, 178, 225
 mean solar, 31
 standard, 31
 Tokugawa Era, 202, 268, 277
 topology, 8, 13, 17, 162, 440, 456–463, 535,
 537, 540
 algebraic, 448
 combinatorial, 448, 456–457
 differential, 448
 point-set, 448, 461–462, 534
tópos, 456
 torsion, 477
 torus, 118, 458, 479
 total curvature, 470
 Toulouse, 420
 Tours, 318
 Tours, battle of, 283
 town, 257, 261
Tractatus de latitudinibus formarum, 330
 tractrix, 486, 491, 492
*Transactions of the London Philosophical
 Society*, 456
*Transactions of the American Mathematical
 Society*, 491
Transactions of the Royal Irish Academy,
 443
 transcendental extension, 443
 transcendental function, 169
 transcendental number, 445, 522, 536
 transfinite, 534
 transfinite induction, 535
 transformation
 linear, 452
 Möbius, 463
 orthogonal, 455
 transformation groups, 285
 transformation of area, 114, 117, 141, 215
 transformation of volume, 117
 transition matrix, 429
 translation, 318, 452
 transportation, 248
 Transvaal, 409
 transversal, 139, 172
 trapezoid, 22, 49, 69, 251
 curvilinear, 235
Trattato d'algebra, 335
Treatise of Fluxions, 382
Treatise on Large and Small Numbers, 269
Treatise on Optics, 305
Treatise on the Latitude of Forms, 330

- Treatise on the Projective Properties of Figures*, 451
- tree, 253, 257, 261, 352
- Treviso, 328
- triad, 72
- triangle, 6–8, 22, 47, 53, 68–69, 85, 87, 88, 104, 112, 117, 121, 138, 145, 176, 177, 191, 219, 220, 227, 251, 362
- angle sum, 103
 - axial, 119, 162
 - curvilinear, 235
 - equilateral, 123, 279, 316
 - isosceles, 82, 199, 215, 229, 278, 365
 - Pascal's, 218, 434
 - right, 72, 152, 198, 216, 221, 247, 249, 257, 261, 365, 371
 - integer-sided, 172
 - spherical, 308, 317, 332, 484
- triangles
- congruent, 21, 137, 173, 175, 278, 304
 - similar, 224, 261, 317
- triangular number, 91, 95, 109, 316
- trichotomy, 130, 157
- Trigonometriae sive de dimensione triangulorum libri quinque*, 332
- trigonometric function, 6, 184, 308, 323, 378, 500
- trigonometric series, 390, 495, 518, 526
- uniqueness, 532–533
- trigonometric tables, 30
- trigonometry, 53, 205, 209, 219–226, 235, 238, 250, 261, 271, 288, 304, 320, 329, 342, 358, 382, 406, 484, 487, 492
- plane, 290, 308, 320
 - spherical, 184, 290, 305, 308, 320
- Triparty en la science des nombres*, 320, 335
- Tripes Examination, 409
- trisecting the angle, 116, 139, 147, 286, 297, 333, 342
- trisection, 106, 122–125, 305, 444
- trisector, 199
- trochoid, 453
- Troy, 107
- truncated icosahedron, 349
- trunk, 352
- truth tables, 544
- Tschirnhaus transformation, 434–436, 438, 439
- tuning fork, 10
- turbot, 59
- Turkey, 27–29, 88, 287
- Turks, 204
- Tusculan Disputations*, 148
- Twelfth Dynasty, 58
- two mean proportionals, 118, 166, 342
- two-dimensional π , 216, 219, 227, 251
- two-dimensional space, 217
- two-line locus, 167, 193
- Tyre, 177, 179
- Ujjain, 208
- ultima Thule*, 178
- ultra kardinem*, 174
- Umayyad Dynasty, 283–285
- Umayyad Empire, 204
- unbounded, 476, 483
- uncountable set, 538
- undecidability, 555, 556
- unenumerable number, 217, 225
- unicursal graph, 17, 22
- uniform, 330
- uniform convergence, 525–526
- uniformly accelerated motion, 320
- uniformly difform, 330
- unique factorization, 110
- unit, xxv, 6, 30, 55, 179, 184, 195, 242, 249, 484
- unit price, 248
- United Kingdom, 30, 240
- United Nations, 240
- United States, 30, 329, 409, 485
- Universal Arithmetick*, 453
- universe, 208, 211, 241, 547
- universities
- Indian, 210
 - University of Amsterdam, 551
 - University of Berlin, 387
 - University of Bonn, 454
 - University of California, 431
 - University of Erlangen, 412
 - University of Freiburg, 14
 - University of Göttingen, 387, 406, 409, 412, 457
 - University of Geneva, 16
 - University of Hawaii, 54
 - University of Jena, 407
 - University of Kansas, 410
 - University of Kazan', 487, 488
 - University of London, 241, 405
 - University of Madras, 211
 - University of Moscow, 527
 - University of Padua, 321
 - University of Pavia, 482

- University of St Andrews, xxvi, 229, 409
 University of Wisconsin, 149
Upanishads, 204
 Upper Egypt, 56
 Ur, 28
 urn model, 422
 utility, 424
 Uzbekistan, 287, 289
- Valmiki Ramayana*, 213
 Vandermonde determinant, 442
 vanishing point, 352
 variable, 311, 347, 359, 383, 495
 complex, 401, 500
 dependent, 514
 independent, 514
 random, 428
 real, 500
 variance, 425
 variation, 394
 second, 394
 variety, 460
 closed, 460
 VAT 8402, 45
 Vatican Library, 196
 vector, 452, 474
 vector space, 4, 452
 vectors, orthogonal, 474
Vedas, 201, 204, 205, 213–218
 Vedic mathematics, 205
 Vega, 488
 velocity, 320, 330, 374
 instantaneous, 375
 Venus, 181
 vernal equinox, 180, 225
 verse, 233
 versed sine, 275
versiera, 399
 vertex, 173, 184
 Věstonice wolf bone, 15
 vibrating membrane, 394, 476
 vibrating string, 394, 400, 512–513
 vicious circle, 109
 Vienna, 320, 406, 414
 Viet Nam, 239
 vigesimal system, 31
 Vigevano, 322
Vija Ganita, 209, 233
vikalpa, 217
 Vikings, 315
 vine, 253
 virtual certainty, 422
- volume, xxv, 6, 46, 49, 58, 73–76, 80, 150, 169,
 193, 195, 219, 238, 243, 251, 254, 262,
 271, 274, 328
 “neat”, 228
 “practical”, 228
 “rough”, 228
 Vorderasiatisches Museum, 43
Vorstudien zur Topologie, 456
 Vulgate, 81
vyayam, 214
- WA, 268
 Walla Walla County, 466
wasan, 202, 268–270, 280, 289
 water, 84
 wave equation, 390, 512, 513, 519
 wave function, 427
 wave, sine, 10
 weak extremal, 395, 397
 weather forecasting, 418
 weaving, 17
 Weber–Fechner law, 425
 wedge product, 474
 Weierstrass approximation theorem, 528, 529
 Weierstrass M -test, 526
 weight, 6, 30, 129, 242
 well-defined, 145
 well-ordered, 537
 well-formed formula, 555
 wheat, 256, 265
 Whittaker–Shannon theorem, 11
 width, 19, 41, 42, 69, 142, 249
 wine, 328
 witch of Agnesi, 399
 wolf bone, 15
 women mathematicians, xxiv, 405–416
 Woolsthorpe, 373
 World War I, 411, 412
 World War II, 204, 240, 411
 World’s Columbian Exposition, 410
 Worldwide Web, 149, 461
- xian*, 221, 249
Xiangjie Jiuzhang Suan Fa, 243
Xugu Suanjing, 257
- Yale Babylonian Collection, 29, 36, 48, 109
yang, 255
Yang Hui Suan Fa, 243
Yang Hui’s Computational Methods, 243
 Yangtze River, 250
 yard, 30

- YBC 11120, 48
YBC 4652, 29
YBC 7289, 36, 39, 109
YBC 7302, 48
year, 12, 30, 180, 187, 230
Yellow River, 250
yen, 274
yenri, 274
ying, 255
Yuan Dynasty, 240
- Zaire, 17
zenith, 224, 285
- Zeno, 461
Zermelo's axiom, 537
zero, 33, 59, 213, 216, 232, 245, 323, 378
 division by, 234, 235
zetetics, 341
zhang, 252
Zhou Bi Suan Jing, 201, 241, 249–251, 253
Zhou Dynasty, 239, 241
Zhui Shu, 243
zij, 288
zoology, 239
Zürich, 413
Zürich Polytechnikum, 523