Imre Lakatos and the Guises of Reason

John Kadvany

Imre Lakatos and the Guises of Reason

Science and Cultural Theory

A series edited by Barbara Herrnstein Smith

and E. Roy Weintraub

Imre Lakatos and the Guises of Reason

John Kadvany

Duke University Press   Durham and London 2001

Printed in the United States of America on acid-free paper ∞
Typeset in Trump Mediaeval by Keystone Typesetting, Inc.
Library of Congress Cataloging-in-Publication Data appear on the last printed page of this book.

Portions of chapter 1 appeared in "A Mathematical Bildungsroman," *History and Theory* 28, no. 1 (1989): 25–42. Portions of chapter 2 appeared in "The Mathematical Present as History," *The Philosophical Forum* 26, no. 4 (1995): 263–87. Portions of chapter 4 appeared in "Reflections on the Legacy of Kurt Gödel: Mathematics, Skepticism, Postmodernism," *The Philosophical Forum* 20, no. 3 (1989): 161–81. Portions of chapter 6 appeared in "Reason in History: Paul Feyerabend's Autobiography," *Inquiry: An Interdisciplinary Journal of Philosophy* 39, no. 1 (1996): 141–46. Portions of chapter 12 appeared in "Verso and Recto: An Essay on Criticism and Social Change," *Cultural Critique* 1, no. 1 (1985): 183–215.

Something I owe to the soil that grew—
More to the life that fed—
But most to Allah, Who gave me two,
Separate sides to my head.

—RUDYARD KIPLING, *Kim*

The genuine refutation must penetrate the opponent's stronghold and meet him on his own ground; no advantage is gained by attacking him somewhere else and defeating him where he is not.

—G. W. F. HEGEL on Spinoza, *Science of Logic*

This page intentionally left blank

For Elena, Daniel, and Susan,

who continue to delight and enlighten

me every day.

This page intentionally left blank

Contents

Analytic Contents xi
Preface xvii
Introduction: Who Was Imre Lakatos? *Ki volt Lakatos Imre?* 1

I | A Mathematical Bildungsroman 21

1 The Mathematical Present as History 23
2 The Method of Proofs and Refutations 45
3 Mathematical Skepticism 85
4 Between Formal and Informal 111
5 Reason Inverted 131

II | A Changing Logic of Scientific Discovery 145

6 Kuhn, Popper, Feyerabend, Lakatos 147
7 A Historiographical Toolkit 157
8 Contradiction and Hindsight 189
9 Reason in History 213
10 A Changing Logic 227
11 Classical Political Economy as a Research Programme 237

III | Magyarország/Hungary 263

12 Hungary 1956 and the Inverted World 265

Notes 317
Bibliography 359
Index 371

This page intentionally left blank

Analytic Contents

Magyar: Hungarian, Magyar

magyarázni: to explain

Introduction: Who Was Imre Lakatos? Ki volt Lakatos Imre?
There are two Lakatoses. One is the outstanding successor of Karl Popper, editor of the *British Journal for the Philosophy of Science* at the time of his death and key participant in the Anglo-American philosophy of science debates of the 1960s and 1970s. The other is a Hegelian who covertly introduced innovative ideas about history, reason, and criticism into Anglo-American philosophy. This combined dual philosophy makes reading Lakatos analogous to the 1920s' reading by Georg Lukács' in *History and Class Consciousness* of the buried yet substantive Hegelian architectonic in Karl Marx. The problem here is not that of understanding Lakatos but rather his marvelous philosophical, historical, and cross-cultural achievements. Several themes in Lakatos's work suggest an interpretation of his dual philosophy against the perverse irrationalism of Hungarian Stalinism between World War II and the emergence of the failed 1956 Hungarian Revolution.

Part 1 A Mathematical Bildungsroman

1. The Mathematical Present as History
Lakatos's *Proofs and Refutations* is an unusual combination of mathematical history and an analysis of numerous nineteenth-century proofs of Leonhard Euler's theorem about polyhedra. Lakatos metaphorically describes his complex narrative technique using Ernst Haeckel's "biogenetic law" that "ontogeny recapitulates phylogeny," also a shorthand for Hegel's historiographical technique in *The Phenomenology of Spirit.* The strange historiography is Lakatos's means for devising a historical and fallible account of modern techniques of mathematical proof. He identifies formalism—meaning the complete identification of mathematics with some formalized, metamathematical representative—as the philosophical perspective he intends to challenge through his history.

Lakatos develops a theory of mathematical heuristic, or a proof pedagogy, through a history of nineteenth-century mathematics and contemporary formalism, analogous to Hegel's approach to constructing an exoteric and learnable philosophical method by summarizing historical patterns of knowledge. The bildungsroman genre, created by Johann Wolfgang von Goethe in his *Wilhelm Meister* novels, contains several of the main narrative techniques used both in the *Phenomenology* and *Proofs and Refutations* to present rationality as a process of self-formative learning through error. The "hero" of *Proofs and Refutations* is Euler's theorem, whose historical odyssey is analogous to that of the generic learning consciousness in the *Phenomenology* and Wilhelm's *Lehrjahre.* Hegel's polemical attack against the antipedagogical and aristocratic epistemology he found in Friedrich Schelling's intellectual intuition, against which the *Phenomenology* is positioned, will be repeated by Lakatos in his critique of the Euclidean-deductivist style in mathematical texts. Schelling's irrationalism will provide a route later back to Lakatos's Hungary via Lukács's views on Hegel and the origins of modern irrationalism described in Lukács's 1954 *The Destruction of Reason.*

2. The Method of Proofs and Refutations

Theorems for Lakatos are conjectures, and an informal proof is a thought experiment or decomposition of a mathematical conjecture into subconjectures, then organized as a series of lemmas that explain the truth of the conjecture. As well as much else, *Proofs and Refutations* explores the relationship between the improvement of informal proofs in history and traditional logical justification, truth, and falsity. The technique Lakatos calls "lemma-incorporation," reformulated as the method of proofs and refutations, is the engine for these interactions. For Lakatos, the method of proofs and refutations is a historical innovation of mid-nineteenth-century mathematics through which rigor is improved as theorems become wedded to the mathematical objects they are about. His account of the coupling between mathematical objects and the theorems describing them is a version of Hegel's so-called phenomenological criticism, being the historiographical method of the *Phenomenology.* Lakatos's history of the method of proofs and refutations gives a reasonable portrayal of major methodological problems raised by nineteenth-century mathematics, culminating with Georg Cantor's development of point set theory through his work on trigonometric series.

3. Mathematical Skepticism

Lakatos attributes the foundational crisis in mathematics at the start of the twentieth century, in part, to the skeptical power inherent in the method of proofs and refutations. This is consistent with Hegel's phenomenological method as a historiographical reformulation of central Pyrrhonian skeptical tropes. The principal differences between classical Pyrrhonism and Hegel's and Lakatos's transmogrifications are the goals to which skeptical criticism leads. Lakatos explains mathematical fallibilism as a species of mitigated skepticism in which its radical and destructive side is tempered for the sake of mathematical progress.

4. Between Formal and Informal

Lakatos expresses the contemporary relationship between formal and informal proofs as a problem of fallible translation. The heuristics underlying Kurt Gödel's incompleteness theorems, considered as informal mathematical theorems, illustrate the translation process and skepticism in modern mathematics. A history of "monster-barring" and lemma incorporation for Gödel's second incompleteness theorem supports Lakatos's ideas on translation within metamathematics as informal mathematical practice. Gödel's proof, therefore, proves that it is itself a piece of informal and historical mathematics.

5. Reason Inverted

Lakatos criticizes the Euclidean-deductivist style found in many mathematics texts as concealing the heuristic logic underlying their proofs. Appeals to mathematical intuition found in such accounts play the role of Schelling's elitist intuitionism, which Hegel wanted to prevent through his historical pedagogy. In this way, the Euclidean-deductivist style promotes a specifically Hegelian species of irrationalism. The ancient Greek method of analysis-synthesis and the history of concealed proofs both provide precedents for the means by which the heuristic logic of proofs and refutations is concealed in traditional deductive proofs. Marx described a similar inversion of historical development and logical exposition for his economic historiography. The foundation for articulated history and heuristic, as the opposite of inarticulate and irrationalist intuitions as a source of knowledge, is the autonomy of natural language as adopted by Lakatos through Popper's "objective knowledge," but is also found in Hegel's conception of language as the essential medium for cultural *Bildung.*

Part 2 A Changing Logic of Scientific Discovery

6. Kuhn, Popper, Feyerabend, Lakatos

A principal lesson of the philosophy of science debates of the 1960s and 1970s is the central role for history in understanding scientific method.

Lakatos makes historiographical problems his main philosophical theme, combined with strong normative claims about how the criticism and appraisal of scientific theories ought to proceed.

7. A Historiographical Toolkit

Rather than using a single theory confronting observations as his unit of analysis, Lakatos uses series of theories changing in time and placed within the larger context of competing research programmes. Applications of the methodology of scientific research programmes by Lakatos and his students are summarized, including accounts of the early wave theory of light, oxygen and phlogiston, the phenomenological and kinetic theories of heat, Copernican and Ptolemaic astronomy, and the special theory of relativity. The latter two studies show the problems faced by the methodology of scientific research programmes in explaining truly revolutionary science.

8. Contradiction and Hindsight

Lakatos's two strongest criticisms of Popper are those associated with the role of contradictions in scientific theories and the difficulties in specifying crucial falsifying experiments in advance of conducting them and interpreting the results. Lakatos argues that contradictions are ubiquitous and do not by themselves provide grounds for theory rejection. Theory rejection also occurs often with historical hindsight and the reinterpretation of experiments provided by theories not available when the experiments were conducted. Niels Bohr's early quantum theory provides Lakatos's main evidence for his claims about contradictions. The Michelson-Morley experiments on the ether and speed of light, and their role in the genesis of relativity theory, provide support for the role of hindsight.

9. Reason in History

After developing his theory of research programmes as a critique of Popper and applying it through several case studies, Lakatos makes the topic of historical reconstruction explicit. Philosophies of science become historiographical theories by characterizing just what in history was science. Likewise, any history of science contains an implicit normative methodology of how science proceeded. All history of science is theory laden by some methodological theory to define its "internal" core abstracted from who created it and under what material or social conditions. "External" history is used to explain the actual historical conditions through which the internal content appeared. Lakatos's principal historiographical distinctions and claims about the theory ladenness of historical inquiry occur explicitly in Hegel's historiographical discussions.

10. A Changing Logic
Lakatos creates his historiographical theory via a self-application of the theory of research programmes to its own emergence in the philosophy of science. A similar skeptical argument was deployed by Lukács in *History and Class Consciousness* to argue for the changing historical status of Marxist theory. With Lakatos, self-application also leads to a changing logic of scientific discovery and completes his Hegelian vision. Lakatos's historiographical theory, moreover, turns his descriptions of scientific progress into normative tools for characterizing scientific progress and degeneration.

11. Classical Political Economy as a Research Programme
The long-standing question of the scientific status of Marxian economics is answered in the affirmative for the research programme that Marx conceived his work to belong to—namely, the tradition of classical political economy, and the work of Adam Smith and David Ricardo. Marxian economics does not grow into a progressive research programme but Marxian economics nonetheless qualifies as a research programme and hence is "scientific." Thus, Lakatos implicitly settles the problem of reconciling the two Marxes: one the historicizing social theorist and son of Hegel; the other the nineteenth-century contributor to scientific political economy.

Part 3 Magyarország/Hungary

12. Hungary 1956 and the Inverted World
From the end of World War II until 1956, Hungarians lived a veritable life of lies in which dissemblance, surveillance, paranoia, fear, and the falsification of political life and history permeated social interactions and cultural life. The essence of Stalinism was the personality cult, centered in Hungary on Mátyás Rákosi. Many writers and other leading intellectuals became elite Stalinists, contributing their skills to perpetuating Hungary's domination from within by Soviet rule. Ritual practices of self-criticism and pseudo-dialectics completed the perversion of Hegelian-Marxian reason. Stalin's death in 1953 and the release of thousands of political prisoners led to a slow awakening and turnabout of Stalinist intellectuals against the Rákosi regime. The standard history of 1956 views the revolution as based in demands

for the right to truth, meaning not having to lead a dissembled life in a society that had effectively lost workable conceptions of truth and falsity.

Lakatos was a charismatic and treacherous member of the Communist underground during World War II, after which he worked and studied with Lukács. A political operative in league with some of the highest members of the Stalinist government, Lakatos wrote ideological criticism and conducted disruptive infiltration work at Budapest's distinguished Eötvös College, helping to bring about its downfall. Yet, caught up in Rákosi's purges, Lakatos was imprisoned during Hungary's "Ice Age" in the infamous Recsk forced labor camp. After his release, he eventually made an about-face like other Stalinist intellectuals. His speeches at the important Petőfi Circle debates in the fall of 1956 on educational topics anticipates his later views on scientific criticism and objectivity. Lakatos was a classic Hungarian Stalinist intellectual of the postwar era.

Lukács's *The Destruction of Reason* contains an implicit critique of the Stalinist personality cult via the book's account of Hegel's attack on Schelling's intuitionism. Shortly before he died, Lakatos parroted Lukács's ideas in polemical attacks, suggesting comparisons between his own conceptions of irrationalism and Stalinist Hungary. The latter appears as a ghastly irrationalist perversion both of Hegelian-Marxism and Lakatos's historicist rationality. Lakatos's philosophy provides a window into this macabre world, warning of the power and dangers manifested by several guises of reason.

Preface

As I will say in the text proper, this book is not an intellectual biography. It is an expository history of ideas, including some political history, motivated by an apparent covert philosophy embedded in the English-language works of Imre Lakatos. If you want, the book is a hidden or secret history—not of subterfuges or intrigues but of ideas and histories temporarily separated that their progeny might survive and prosper. An intellectual biography of Imre Lakatos would be of great interest, although I conjecture that the many ambiguities of Lakatos's work and the betrayals of his life entail that we learn most of philosophical value by not attending too much to *him.*

Guises of Reason's thesis is that Lakatos's English-language philosophy of science and mathematics is a philosophical palimpsest, containing an original and instructive account of historical rationality deriving from Hegel, Marx, and the Hegelian-Marxism of one of Lakatos's teachers in Hungary, Georg Lukács. Hegel is the historicist philosopher par excellence, so Lakatos's creation, whether intended or not, makes available a rich tradition involving historiographical method, educational philosophy, and criticism, now set in the context of a contemporary philosophy of science and mathematics. Lakatos's Hegelianism is neither eclectic nor opportunistic. It is an inspired and powerful vision of scientific and mathematical rationality as historical processes. This vision is also developed sub rosa within Anglo-American philosophy, so that Lakatos's work lives in two philosophical worlds. Like no one else, Lakatos married substantive philosophical ideas from Anglo-American and continental philosophical traditions, and made them one. This Janus-faced philosophy, as we will see, becomes an entry point into Lakatos's terrifying and morally compromised Hungarian past as well.

Hegel has repeatedly been proclaimed a dead philosopher every several decades over the last two centuries, only to be rehabilitated in some novel incarnation. The most important reinvention during the twentieth century was Lukács's convincing interpretation of the positive role for human consciousness and subjectivity in Marx's social theory, proposed in opposition to the crude mechanical Marxism popular since Marx's death and promulgated often in the past as political orthodoxy. Lakatos provides a marvelous continuation of the Hegelian heritage by using Hegel to historicize the philosophy of mathematics and science. While Lakatos's work leads to politics, and especially the despised Stalinist politics of his native Hungary, his role for history is not to politicize science. Lakatos instead makes history central to scientific reason while trying to harbor scientific criticism from political bias. Exploring this laudable and ambitious project, and unraveling some of its contradictions, is one of this book's several goals.

Readers of the book in manuscript asked me, "But what ultimately do you think, was Lakatos consciously or unconsciously a Hegelian?" The best answer is that equally strong arguments can be made for each side. That is, there is sufficient evidence to justify that Lakatos was either consciously dissembling his Hegelianism or merely integrating his past and present with little special intent.

To give an example of how that equipollence for the two possibilities might go, a piece of Lakatos's biography is helpful. He was not a deep reader of Hegel and learned much of what he knew about Hegel only from Marx and other sources, including a close relationship with the Hungarian Marxist philosopher Lukács. Indeed, many of the key ideas discussed in this book can be found to have versions outside of Hegel that a quick study, like Lakatos, could absorb and turn to his own use, almost ignorant of their provenance. There is then the prima facie case that Lakatos was a voracious and even somewhat indiscriminate consumer of ideas, and the appearance of all kinds of traces of Hegel and others in his work expresses no more than that.

But just which Hegelian ideas are found in Lakatos? Not a miscellany but several of Lukács's distinctive and appealing views of Hegel as philosopher-pedagogue, antiromantic, and consummate methodologist. With a bit more attention, we find also in Lakatos adroit reformulations of some of Hegel's aphorisms and historiographical concepts—not just as ornament but as summarizing some of Lakatos's most creative thinking. Mere coincidence? Just picked up en passant from Lukács or elsewhere? Perhaps. Then again, there is enough of a trail to suggest considerable learning behind the appearance.

All one can do then, in the face of such strong arguments for both sides of the question, is to suspend judgment on Lakatos's true intent. Of course, that is where a successful dissembling would lead as well.

Paul Feyerabend, another Central European émigré philosopher, and Lakatos's close friend and most acute critic, occasionally alluded to Lakatos's apparent subversion of Anglo-American philosophy from within, and now that story is told. Feyerabend was a thoroughgoing skeptic about the validity of universal rules for scientific progress and was largely correct in noting the difficulties in accounting for many great scientific achievements even using Lakatos's robust historicism. On one level, then, Feyerabend had the last word in their powerful and ongoing debate over scientific reason. But now the intellectual duel between these two late master samurai takes a surprising turn. Feyerabend was also keen to point out that scientific progress was *sometimes* critically aided by rhetoric, polemic, or subtle shifts of meaning or argument, including the clever insinuation of concealed ideas—not as dishonesty but as practical means of bringing about changes in thinking patterns, to try out and explore new ideas in inhospitable intellectual environments. Now who did that, if not Lakatos? The philosophical heritage of these two children of twentieth-century catastrophes—Feyerabend, though not a Nazi, was an Austrian army officer in World War II crippled for life by shrapnel, and like Lakatos, made his career much through English-language philosophy—includes a shared appreciation for the importance of political tactics of all kinds in intellectual battles, whether philosophical or scientific. Surprising, though, is that it would be Lakatos, the champion of rationalism against Feyerabend, who would turn out to be an exquisite philosophical example of just the covert techniques Feyerabend found in the history of science.

Via Lakatos, Feyerabend, and Hegel, much of *Guises of Reason* is informed by the history of skepticism. Yet skeptical methodology does not imply an amoral outlook. For the purposes of this book, in distancing ourselves from trying to decide Lakatos's truth, or understanding him, we notice similar and strange parallel problems of truth and falsity in the Stalinist Hungary that Lakatos left behind for England in 1956. "A whole nation cannot be suspect," János Kádár, Hungary's postrevolution Soviet puppet, said in 1964. Unfortunately, Kádár was not describing Hungary; he was making a normative appeal for how life *should not* be, a plea to extricate Hungary from its pathological culture of uncertainty and fear. Lakatos's philosophical work provides, along with much else, a window into that world, and the perverse relations of reason and terror that helped it function. In caricature, Lakatos's Stalinist Hungary was itself predicated on patterns of skeptical undecidability created through dissemblance, surveillance, informing, and betrayal. In broad epistemic terms, the depravity of Stalinism was an evil, inverted twin of skeptical, even scientific, criticism. In this way, the fascinating contradictions of Lakatos's many-layered ideas and life are those of the dangerous world he fled.

And while the 1956 Hungarian Revolution was crushed, the revolt was nonetheless ingeniously created from within Stalinist irrationalism using the powers of reason, courage, language, solidarity, and hope against darkness. Reason and skepticism are intimately related, but political and intellectual progress are possible through choice and sometimes heroism.

I would like to provide a cheerier and more secure account of Lakatos's world, but that would be false to his work, the history in which he was caught up, and the histories he created as a resourceful political operative in Hungary and, later, world-famous philosopher in England. Imre Lakatos was a survivor whose philosophy is a testament to the cunning survival of reason in several of its guises.

Finally, some guidance on reading this book, which combines ideas from the history and philosophy of science and mathematics, Hegel, nineteenth-century political economy, and Hungarian political history. To make the case that Lakatos's Hegelianism is of intrinsic interest, regardless of his intentions, meant providing sufficient detail on his several worlds. Table 1 on page 294 lists the major Hegelian themes in Lakatos's work and includes chapter references. For ease of reading, chapter 2 on the method of proofs and refutations as well as chapter 4 on Gödel's proofs could be skimmed on a first pass. A brief appendix at the end of chapter 2 introduces many of the mathematical ideas discussed. Readers so inclined are invited to start with Lakatos's philosophy of science in chapters 6 through 11, or even with chapter 12 on Hungarian Stalinism and Lakatos's nefarious past. Those chapters are sufficiently self-contained so that, like this preface, they provide alternative beginnings as well as endings.

Several friends and colleagues offered encouragement, advice, or comments during the writing of this book, including Harvey Weiss, Jerry Ravetz, George Marcus, Jancis Long, László Ropolyi, Gábor Kutrovátz, Martin Jay, Luce Giard, Robert Paul Wolff, Thomas Trebitsch, Brit Katzen, Reuben Hersh, and Reynolds Smith of Duke University Press; Wendy Lesser published my first essay on Lakatos in *The Threepenny Review;* Cindy Milstein was a superb copyeditor; and Gábor Forrai generously read and commented on the entire manuscript. *Köszönöm szépen.* A special thanks goes to my parents, who left Budapest in 1946.

Introduction: Who Was Imre Lakatos? *Ki volt Lakatos Imre?*

Two souls, alas, are dwelling in my breast

And one is striving to forsake its brother.

—GOETHE, *Faust*

Imre Lakatos, who died in 1974 at the age of fifty-one, was one of the most original philosophers of science of the twentieth century. Though not as well known outside philosophy as his contemporaries Thomas Kuhn and Paul Feyerabend, or his mentor Karl Popper, Lakatos was a central figure in the philosophy and history of science debates of the 1960s and 1970s that created a sea change in received views of scientific method and practices throughout the humanities, social sciences, and science itself. Lakatos was Popper's most assiduous disciple and greatest critic. He surprisingly extended Popper's idea that scientific theories can never be proven to mathematics in his masterpiece *Proofs and Refutations: The Logic of Mathematical Discovery,* originally part of his 1961 Ph.D. thesis at Cambridge, where Lakatos had arrived as a superannuated graduate student after fleeing Hungary and the failed 1956 Hungarian Revolution. Lakatos's philosophy of science similarly broadened Popper's view that science is marked not by foundational truths but by "conjectures and refutations," or the creation of scientific progress by the continual proposal and elimination of successively improved theories. Theories are never proven true in Popper's view, but they can be falsified, and progress through such falsification is the key idea of Popper's *The Logic of Scientific Discovery,* published originally in Vienna in 1934 as *Logik der Forschung.* Although in *Proofs* Lakatos appears to seamlessly continue Popper's notions, in the philosophy of science Lakatos ravaged Popper's philosophy with incisive criticism even as he built on it, earning his illustrious teacher's enmity as his own fame grew. The son turned against the father, and the father recognized the wicked turn against him even as it fulfilled the spirit of his very own critical rationalism. Twenty-five years after his death, Lakatos is still being read in several languages. *Criticism and the Growth of Knowledge,* the 1970 volume Lakatos edited with Alan Musgrave at the height of the debates over Kuhn's *Structure of Scientific Revolutions*—itself one of the most important postwar works of English-language philosophy—

has gone through over twenty printings, and is a milestone in the rejection of the asocial and antihistorical legacies of logical positivism.

A fair statement of the received view of Lakatos's philosophy is that while giving the same central role for history as did both Kuhn and Feyerabend, his is a more sophisticated version of Popper's ideas, perhaps with greater philosophical nuance than Kuhn yet less methodological radicalism than Feyerabend. This picture of Lakatos as occupying a judicious middle ground is incomplete, creating a perception of less radical and far-reaching ideas. Between his arrival in England in 1957 and his death there in 1974, Lakatos developed not one philosophy but two, both of which are contained in his always eloquent, scholarly, and entertaining essays. One philosophy is overt, that of Popper's best son who did him one better.[1] The second is covert, or at least not obvious to several generations of readers. This latter philosophy is a brilliant and sui generis Hegelian-Marxism, ideologically and philosophically light-years away from Popper and most members of the Anglo-American philosophical and social scientific circles making up Lakatos's primary readership. Imre Lakatos was a minor philosophical genius whose strange, compelling, and profound essence was that his works appear as the dissembling of a covert Hegelian taking the Popperian castle from within.

The purpose of this book is twofold: to substantiate this second Hegelian reading of Lakatos's philosophy of mathematics and science; and to take stock of Lakatos's Janus-faced achievement for contemporary philosophy, the history of ideas, and his native Hungary. With the fall of the Berlin Wall, and the political transformation and renormalization of life in Hungary, it is finally appropriate to raise questions about Lakatos's work that take the passage of time and catharsis—political, philosophical, intellectual, and for some personal—to approach with calm and reason.

As mentioned in the preface, Lakatos's work ultimately leads back to the treacherous times preceding the 1956 Hungarian Revolution. This passage into Lakatos's history is the most problematic and crucial step in grasping his complex world. For if Lakatos's work is indeed two philosophical worlds as one, it would be unsatisfactory to raise basic questions about his past only to leave them unanswered. Thanks to the opening of Hungary's archives, we now have a detailed account of Lakatos's Hungarian past, which can be used to complete our understanding of his dual philosophy. It is likely that what is now known about Lakatos is as much as will ever be known about him, or is at least as much as needs to be known in order to assess his place in the ideas and politics of the strange and uncomfortable century just ended.

Lakatos's biography coheres quite well with numerous themes in his work. But as noted earlier, this book is not about Lakatos the person, or is about him very little. Rather, it is about the network of ideas in his philosophical work and a history of ideas relevant to them. The abbreviated history of the Hungarian Revolution with which the book concludes, while including Lakatos's role, is presented here as an object lesson in a type of irrationalism that is intimately related historically and conceptually to much in Lakatos's work. One of Popper and Lakatos's notions about knowledge was that language provided the means for ideas to become semiautonomous from their producers, allowing others, as Popper implied, to kill my ideas but not me.[2] So, too, an analysis of Lakatos's intellectual kinships and genealogies, and their relationship to ideas appearing in Stalinist Hungary, stand on their own regardless of his intentions. There is no need to appeal in an essential way to what Lakatos might have read, known, or meant. Just as Lakatos had a bias for histories excluding social or psychological speculation, likewise, no such explanations are necessary in reading Lakatos's work, interesting as they might be. Many are indeed now wondering, "*Ki volt Lakatos Imre?* Who Was Imre Lakatos?" as asked in Hungarian and English by the title of a Hungarian documentary on Lakatos's life. Naturally, our interest in Hungarian history is prompted by knowing that Lakatos came from Hungary, and was there an active intellectual and Communist. But his history in particular plays little role in the conclusions here: that the political and social world of Stalinist Hungary was a culture in which the abstractions of Lakatos's philosophical work vibrantly came to life, albeit in an inverted and perverted shape.

In this historical reality, several features of Hegel and Lakatos's historical rationalism become thoroughly inverted into their irrationalist doubles. The double character of Lakatos's texts and ideas is his legacy for us, even as he drops out and leaves the problems of how we might nonetheless relate this game of doubles to some concrete and meaningful history. It is rather part of the beauty of Lakatos's work that, without being about Lakatos per se, Hungary 1956 provides a solution to the enigma of Imre Lakatos because his writings turn out to be particularly apt to the role of historical reason during the dark times of Stalinist power. As Hegel wrote in his *Philosophy of Right,* "Philosophy too is its own time apprehended in thoughts. It is just as absurd to fancy that a philosophy can transcend its contemporary world as it is to fancy that an individual can overleap his own age, jump over Rhodes."[3] What happens, though, when there are two such worlds—Anglo-American and Hegelian philosophy—each seen imperfectly through the other? Then it is the task of the Hegelian philosopher, living in the former after having been educated in the latter, to make the two worlds one through a philosophy that enables his readers to travel between them. *Hic Rhodus, hic saltus:* here is Rhodes, jump over it if you can.

Those who should profit through this account include philosophers of science and mathematics who could never quite integrate the diversity of Lakatos's views on history and scientific or mathematical method; critical theorists and students of Hegel or Marx, who will find in Lakatos major contributions to explaining scientific criticism vis-à-vis historicism and Hegelian-Marxism;[4] intellectual historians, for the convoluted communication of ideas by a charming and wicked Central European émigré; historians of ideas, for the new light Lakatos sheds on the transmission of skepticism through German philosophy, his own historiographical theory, and his account of nineteenth-century mathematics; and finally, Hungarians and students of the Hungarian Revolution, for Lakatos's implicit rendering of Hungary's perverse Stalinism and the role of compromised Stalinist intellectuals, of which Lakatos is a paradigm case, and their heroic role in the 1956 revolt. Those contributions are listed in no special order: all jointly constitute the main pieces of his intellectual legacy, which should delight and enlighten us all.

The Problem

Where can such a reconstructed past begin? What could lead us to such fantastic claims about the work of a distinguished scholar, now dead so many years?

Scattered throughout Lakatos's philosophical works are several tantalizing reminders of Hegel and Marx that leave one with a sense of déjà vu. For example, in explaining the backward-looking, historicized nature of his "scientific research programmes," Lakatos alludes to the "owl of Minerva," Hegel's famous expression from the *Philosophy of Right* describing philosophy's necessarily retrospective historicism. There is also Lakatos's advice that his caricatured rational reconstructions of the history of science and mathematics, themselves reminiscent of the caricatured histories of philosophy and economics found in Hegel and Marx, can be taken as histories of how science "ought" to have behaved, providing the same normative-critical turn away from Hegel as is seen in Marx. As well, there is Lakatos's own description of *Proofs* in terms of the "ontogeny and phylogeny" of history, a Marxist shorthand for the specialized narrative style of Hegel's *Phenomenology of Spirit.* There is, finally, Lakatos's use of the idea of false consciousness in the history of science, and his criticism of falsified histories of mathematics and science, reminiscent of analogous claims made by Marx against many economists or Marxist politicians against nearly everyone. Many of these Hegelian- or Marxist-sounding concepts are found in an elaborate footnote apparatus that, Lakatos says, makes for numerous "cross-, forwards- and back-self-references, which tend to make my papers difficult to follow. But while being apologetic for this style of exposition, I am unrepentant concerning the content" (*MSE* 212n1). This style is also a clue to Lakatos's Hegelian ways since, as one reader said, it is "as if it were impossible to understand any part of them without understanding the whole."[5] Perhaps the truth is the whole indeed.

Lakatos met Popper in England, where both would teach at the London School of Economics. Sir Karl, writes Lakatos, "changed my life. . . . His philosophy helped me to make a final break with the Hegelian outlook which I had held for nearly twenty years" (*MSRP* 139). Lakatos's apparent rejection of his Hegelian past has frequently led people into believing that for Lakatos, the dialectic was no more than a superficial stylistic ornament. But Marx was once a Young Hegelian as well, and one cannot understand Marx except as simultaneously working out of the traditions of German idealism and classical political economy. "The reason why I continue to criticize various aspects of Popper's philosophy," Lakatos explains, "is my conviction that it represents the most advanced philosophy of our time, and that philosophical progress can only be based—even if 'dialectically'—on its achievements" (*MSRP* 139n2).[6] In the acknowledgments to his Cambridge doctoral thesis, much of which later became *Proofs,* Lakatos remarks that "the three major—apparently quite incompatible—'ideological' sources of the thesis are Pólya's mathematical heuristic, Hegel's dialectic and Popper's critical philosophy" (*MSE* 70).[7]

At first blush, *Proofs* reflects a somewhat vulgar Hegelianism, consistent with even a mechanical Marxist philosophy. For example, Lakatos briefly uses the "thesis-antithesis-synthesis" locution that Hegel never does.[8] In Lakatos's lengthy bibliographies there is also no reference to any work of Hegel or Marx, and at the time of his death no books by Hegel were found in his library. So it might seem that from the beginning, Hegel was little more than a dead dog for Lakatos, that after he joined up with Popper, Lakatos had only coquetted with Hegel's mode of expression. Certainly, if Lakatos's casual and sometimes imprecise remarks about Hegel were all we had to rely on, the merits of wringing from his work some vague and overinterpreted reading would not be worth the effort. But all this also might make us suspicious given that one of Lakatos's primary interests is the relation between history, its rational reconstructions, and its outright falsifications. Would such a meticulous historian not refer to *any* of the works of one of his "ideological influences"? Perhaps Lakatos's pronunciamentos are red herrings, manifestations of a feigned superficial Hegelianism at odds with his intrinsic practice. Indeed, Lakatos's explicit asides about Hegel divert attention from the substantive Hegelianism built into his philosophy and its historical approach to science and mathematics.

Not everyone was misled by Lakatos's facade and his apparent renouncement of Hegel. "Lakatos was a closet-Hegelian. . . . [H]e knew he was one, but didn't want anyone else to," writes Marx Wartofsky.[9] Or in Feyerabend's words: "The only philosopher who secretly imbibes the forbidden brew of Leninism is Lakatos—and the results are evident in his magnificent work. All that is required now is that he confess his vices openly so that others may learn to delight and enlighten us in a similar way."[10] Then there's Ian Hacking, who observes:

> a philosophical émigré may very naturally have a listener on each shoulder, and by dint of unwittingly addressing both, fail to make plain what is being said to either. On the one shoulder is a thoroughly Hegelian and somewhat Hungarian conception of the events of modern philosophy, a body of historical conceptions that Lakatos takes for granted, hardly stating them. On the other shoulder are the English, whose scientific values are just what Lakatos wants, no matter how ignorant and insular the philosophy that runs alongside them.[11]

Yet none of these perceptive readers grasped the full implications of what they glimpsed in Lakatos: a systematic and inventive Hegelian program of great intrinsic interest, not just the "coloring" of a talented refugee's academic art.

It was one of Lakatos's teachers in Hungary, sometime political ally Lukács, who faced a similar conundrum in reading Marx early in the twentieth century. Lukács's spectacular achievement was to recognize both the persistent influence of major Hegelian ideas throughout Marx's mature writings, primarily on Marx's notion of alienation and also the role of human consciousness generally, and Marx's nonmechanical, nondeterministic conception of economic and social history; most important for us is Lukács's repeated insistence on recognizing Marx's methods for historicizing economic and social categories. Though several of the ideas of Hegelian-Marxism were in the air, Lukács more than anyone reformulated the philosophical foundations for Marxism by centering them right on Marx's transformation of Hegel.

History and Class Consciousness, Lukács's seminal contribution to what would become Western Marxism and the critical theory of the Frankfurt School, was all the more remarkable because Lukács did not have the benefit of Marx's then-unpublished *Economic-Philosophical Manuscripts,* which mostly confirmed his conjecture about the underlying Hegelian architectonic of Marx's work.[12] It was indeed Lukács who wrote that for Marx, readers had mistakenly seen Hegel as a "dead dog," and the "dialectic was no more than a superficial stylistic ornament."[13] Our situation with respect to Lakatos's work is similar to that faced by Lukács in unraveling the Hegelian heritage built into Marx's opus. The difference, of course, is that Lukács's work precedes Lakatos's and that the latter was a student of the former. We will also not have any Lakatos-style *Paris Manuscripts* or a *Grundrisse,* the rough draft of *Capital,* replete with Hegelian ideas to be submerged in the final book. But Lakatos was an expert on those processes through which a concept can be decomposed, analyzed, and reconstructed, "thus embedding it," as he notes in *Proofs,* "in a possibly quite distant body of knowledge" (9). Lakatos's own work demonstrates such processes in action while providing a guide to understanding them as well. As Marx once wrote of the unwitting repetitions of political history, "We recognize our old friend, our old mole, who knows so well how to work underground, then suddenly to emerge."[14] Now though, the burrowed dissembling of ideas itself has an explicit philosophical role. That role is the enigma of Imre Lakatos.

Whether Lakatos is really a Marxist or Hegelian is at first a reasonable question, but eventually it self-destructs. Yet regardless of one's views on Marxism, the concealment of Hegel's ideas is a fact in the history of ideas and Marxism, and their discovery is one of Lukács's central achievements. That concealment and discovery, not whether Lakatos purposely hid his Hegelianism, is the relevant historical context for Lakatos's work. For one problem in successfully dissembling ideas is that their origins will be concealed. We can see in Lakatos's philosophy how many Hegelian and sometimes Marxian ideas are transported and creatively used to address issues in the philosophy of science and mathematics, most especially in ways that also address old problems of history and criticism found in Hegel and Marx, and none of it depending on Lakatos's personal attitudes toward history or politics. There is great continuity between Hegel and Lakatos, and even Marx and Lakatos, and innovations in historical philosophy like nothing provided by anyone before Lakatos. Marx's Hegelianism, Lukács showed, was concealed but effectively present in the logic of his social and economic categories. Likewise, Lakatos's philosophy is such a puzzle inside a riddle inside an enigma. In his work, once parallel histories displaced by decades or centuries in time cross continents to intersect and merge. It is no small irony that in trying to understand Lakatos, we are led to Lukács, whose greatest work, *History and Class Consciousness,* involved resurrecting the dead dog Hegel, and whose subsequent career was marked by volte-faces and self-criticisms, both spurious and genuine, in regard to his Hegelianism and politics. These self-criticisms were not unlike those Lakatos offered about Popper freeing him from his Hegelian mind-set.

As for Hegel, pedagogy is a main topic in Lakatos's philosophy, and complements his role as an educator and a pedagogical philosopher. Hegel's various "encyclopedias," for example, are compendiums of learning, or *paideia,* so to speak. From as early as his teens, Hegel had the ambition to influence German culture as a *Volkserzieher,* a people's educator, and his concern with the "education of mankind" directed much of his speculative and historical studies.[15] In addition to creatively reapplying numerous Hegelian methods, the teaching Lakatos provides is a modern education in the historical transformation of ideas. Lakatos's "fascination with research programmes in the history of science reflected his interest in the strategies whereby a good idea might come to power."[16] His own writings are an instance of this process: Lakatos's philosophy is "a splendid Trojan horse," and shows that "if reason is to have a point of contact in this world with its complicated episodes and its hair-raising ideas and institutions, then it must be both sly and sophisticated."[17] "Lakatos was of course a highly talented thinker who contributed many valuable ideas, [but] his chief importance both in Hungary and in England was as an educator promulgating ideas."[18] Part of this lay in Lakatos's Socratic midwifery in leading the reader of his works to those of Hegel and Marx, as well as to other lands where arts of covert rationalism have been practiced as a form of life. One does not always learn what one has set out to learn, as Lakatos says, either because of the subtleties of mathematical proof, or the unstated and perhaps unconscious techniques of a teacher of learning through error. Lakatos combined his role as educator with those of "crusader, journalist, politician," roles that distance Lakatos from Hegel and bring him closer to the post-Hegelian tradition of critical theory from Marx to Jürgen Habermas.[19] Nevertheless, Lakatos is still the Hegelian teacher in that his political bailiwick is the world of ideas, and that is the focus of this book, rather than the controversial university professor and his politics.

As Popper would say, we therefore have *a problem:* Is there a relationship between the philosophical work of Lakatos, Hegel, and Marx, the tradition the latter two initiated, and the philosophy of science and mathematics that

Lakatos developed after leaving Hungary? I conjecture that there is, that it explains and gives substance to a host of ideas in Lakatos's work, clarifies otherwise puzzling arguments between Lakatos and other philosophers of science, and shows Lakatos to have made a fundamental contribution to "the dialectical tradition to which he belongs," from Hegel to Lukács, and even the intellectual history of the 1956 Hungarian Revolution.[20] We may end up with more problems than we begin with, for "a scientific inquiry 'begins and ends with problems.'"[21] In some cases, "the history of a particular problem turns into the history of problems."[22]

Back to Hegel

In *Guises of Reason,* the way will often traverse paths between Lakatos, Hegel, and Popper. Those for Lakatos's philosophy of mathematics and science, here meaning *Proofs* and the methodology of scientific research programmes, intersect sometimes, but overlap little.[23] The philosophy of science does not depend in any fundamental way on the philosophy of mathematics, or vice versa, and each even takes up a very different perspective from Hegel's philosophy. *Proofs* uses history, as Ian Hacking aptly describes it, as a "forward-looking" pedagogical tool to explain how to improve mathematical conjectures into proofs: the model here is the parallel pedagogical summary of philosophical ideas provided in Hegel's masterpiece, *The Phenomenology of Spirit.* Lakatos's philosophy of science, in contrast, is as Hacking says, "backward-looking," using history as a critical tool for the appraisal of scientific theories and not their creative construction; the model here, in part, is the reflexive logic of Hegel's later philosophy.[24] In both the philosophy of mathematics and science, Lakatos's principal achievement was to historicize method using detailed histories of science and mathematics deployed through an innovative historiographical toolkit. The common ground between *Proofs* and Lakatos's "methodology of scientific research programmes"—the British "programme" is retained to refer to this part of Lakatos's work—is Hegelian historiographical technique, deployed in two distinctive Hegelian modes: either the inventive uses of reason to create new ideas, theories, or theorems, or the critical uses of reason in rewriting the past.

The common conceptual foundation between Popper and Hegel that makes Lakatos's simultaneous projects possible is a shared central role for error and learning in human knowledge. The hallmark of science for Popper is not proof, which is in any case unattainable, but criticism and the falsifiability of scientific theories, meaning the potential for refutation by observation or experiment. To Popper's great credit, this rejection of certainty as a goal in scientific knowledge constituted an early and influential criticism of what Albert Einstein called the "fairly dried-up petty-foggery" of logical positivism.[25] Popper's science consisted of bold conjectures and their refutations, and was driven forward by a belief that, as Popper says, "we can learn from our mistakes."[26] Science finds what is good by eliminating the bad, and the

> fundamental difference between my approach and [inductivism] is that I lay stress on *negative arguments* such as negative instances or counter-examples, refutations and attempted refutations—in short, criticism. . . . [A]ll that can possibly be "*positive*" in our scientific knowledge is positive *only* insofar as certain theories are, at a certain moment of time, preferred to others in the light of our critical discussion. . . . Thus even what may be called positive is so only with respect to *negative* methods. . . . It is from our boldest theories, *including those which are erroneous,* that we learn most. Nobody is exempt from making mistakes; the great thing is to learn from them.[27]

The idea of "negative" progress and learning through error is also the starting point for Hegel's introduction to philosophy in the *Phenomenology.* That most forms of knowledge are skeptically subject to error is assumed by Hegel as it is by Popper, and they both share the general idea that progress consists in the improvements to knowledge made by overcoming error. "The most pernicious vice is to seek to preserve oneself from errors," Hegel wrote in a diary a few years before the composition of the *Phenomenology;* "flexibility, educability . . . is truth."[28] Even Lakatos, overstating the case, notes that "Hegel and Popper represent the only fallibilist traditions in modern philosophy" (*P&R* 139n1). Hegel "officially" begins the *Phenomenology* in its introduction, where he rhetorically raises the problem of the nature of cognition: whether it is an instrument by which to gain knowledge or a medium through which knowledge is discovered. What shall it be, in order to begin on the right path? How do we direct ourselves in philosophy so as not to be wrong from the start? To this Hegel responded:

> Meanwhile, if the fear of falling into error sets up a mistrust of Science, which in the absence of such scruples gets on with the work itself, and actually cognizes something, it is hard to see why we should not turn round and mistrust this very mistrust. Should we not be concerned as to whether this fear of error is not just the error itself? Indeed, this fear takes something—a great deal in fact—for granted as truth, supporting its

> scruples and inferences on what is itself in need of prior scrutiny to see if it is true. To be specific, it takes for granted certain ideas about cognition as an instrument and as a medium, and assumes that there is a difference between ourselves and this cognition. Above all, it presupposes that the Absolute stands on one side and cognition on the other . . . it presupposes that cognition which, since it is excluded from the Absolute, is surely outside of the truth as well, is nevertheless true, an assumption whereby what calls itself fear of error reveals itself rather as fear of the truth. (*PhS* 47)

The simple yet great difference from Popper is that Hegel is also pointing out that *our* inquiry into philosophy itself does not escape error either. By implication, the philosophical investigation of knowledge begins with the odd problem of finding its own meaningful starting point and coherent method—that is the purpose of Hegel's introduction to the *Phenomenology.* The following chapter will explore the details of Hegel's learning process, his so-called "phenomenological method," and the historiographical method of the *Phenomenology,* all of which are relevant to the content and historiography of *Proofs.* For now, it is enough that there is a common antifoundational view of knowledge in Hegel and Popper, and that each starts from a shared and quite unabstract idea of learning through error. To be sure, proceed a paragraph forward in either philosopher's work and there is a world of difference: Hegel looked for error in ways of knowing generally, while for Popper, metaphysics and pseudosciences like Marxism and Freudian theory are inferior forms of knowledge because of their insulation from learning; Hegel was used to premodern conceptions of judgment and their quasi-psychological categories, while Popper lives in the twentieth-century world of theories, refutations, and modern logic; Hegel summarized much of the entire romantic epoch and philosophy of his contemporaries, while Popper claims much originality for his work. But a shared starting point of critical philosophy and learning through error commits neither Popper nor Lakatos to sharing the details or even broad outlines of Hegel's philosophy of history, religion, aesthetics, metaphysics, and much else. At the heart of his approach, Popper provides the germ for an innovative Hegelianism focused on mathematical and scientific method as history.

How Hegel and Popper diverge from this starting point is our interest here. For Hegel, in the *Phenomenology,* learning through error almost immediately became an explicitly historical philosophy. The reflexive problem addressed with bravado by Hegel, as suggested by the passage above from the

Phenomenology, is how a philosophy of error and learning itself participates in the process of learning. For Hegel, history and the temporal logic of historical consciousness were the solutions to that problem. Any such approach is completely stopped short by Popper, much as subject-object problems and historical philosophy are generally marginalized in Anglo-American philosophy; that might even define the latter's gestalt. Although Popper said that "scientific theories are perpetually changing" (*LSCD* 71); that "the central problem of epistemology has always been and still is the problem of the growth of knowledge" (ibid. 15); or in quoting Lord Acton, that "there is nothing more necessary to the man of science than its history, and the logic of discovery" (ibid. 14), the importance of history and its significant use remains implicit throughout Popper's philosophy. For Popper, the history of science is a source of examples for a methodology premised on the logical asymmetry of refutation: while one cannot prove scientific theories, contradictions between theory and observation allow theories to be refuted. Popper rejects logic insofar as it cannot be used to prove empirical theories while maintaining logic and the classical law of noncontradiction as the a priori methodological knowledge grounding the logic of scientific discovery. He is adamant that falsification itself is not subject to refutation. In that way, Popper's rejection of positivism turns out to be a halfhearted allegiance. That *some* kind of historicism is implicit in the "negative" progress made via the critical elimination of theories is, for Lakatos, Popper's Achilles' heel. This leaves open the technically sweet philosophical problem of historicizing Popper. That is very much the hidden heuristic underlying the methodology of scientific research programmes. The approach begins simply enough with Popper's falsificationist learning through error and its correlative resistance to history.

What role or credit, then, should we give to Popper in Lakatos's work? In *Proofs,* Popper's primary role is expressed through Lakatos's idea of considering mathematical theorems as fallible and often erroneous conjectures. But Popper has very little influence on how Lakatos addresses any specific historical or philosophical issues, and even the idea of theorems as fallible conjectures draws significantly on ancient Greek thought experiments and the inspiration of Lakatos's university teacher Árpád Szabó. It is in Popper's principal bailiwick, the philosophy of science, and Lakatos's methodology of research programmes that Popper's ideas on contradiction, falsification, and the formulation of crucial experiments figure prominently as targets for Lakatos's systematic critique. As we will see, the specific changes needed to transform Popper's postpositivist philosophy of error into a historicist conception of scientific change all have strong Hegelian affinities. Their coherence can be traced back to Popper or Hegel's antipositivist notion that knowledge has no foundations and changes through the recognition of failure. In rebuilding these ideas either right out of Popper, as in the philosophy of science, or by carrying them along under the umbrella of mathematical history, Lakatos provides an entrée into history that Popper misses.

Learning through error, whether in mathematics or science, is shorthand for a complex of ideas involving contradiction, criticism, and changes in knowledge across time. Error and learning appear in Lakatos in a variety of shapes and guises: through several reformulations of Greek Pyrrhonian skepticism; in methods of scientific criticism; as heuristic methods for discovering and improving mathematical proofs; as conditioning the relation between informal and formalized mathematics; and through techniques for writing history. Popper's own errors in the philosophy of science are also important grist for Lakatos's historicizing mill. As Lakatos presents them, Popper's errors are expressed primarily through his reconstruction of the history of science. The "falsification" of Popper's philosophy sustains Lakatos's own treatment of Hegel's problem of how to overcome the error of the fear of error itself: the conceptual end of Lakatos's philosophy of science is a historiographical theory arrived at through the application of the methodology of scientific research programmes to itself.[29] Rationality is a temporal process for Lakatos, as it is for Hegel, rather than a characteristic of types of beliefs or isolated statements. Lakatos turns Popper's philosophy into a historicist conception of science *and* a theory about the writing of the history of science that acknowledges its own errors in reconstructing the past. The roles for error and learning appearing in Popper are greatly expanded as Lakatos demonstrates in practice. For instance, *Proofs* itself is a series of mathematical proofs applying heuristic methods, and the book itself is overtly about heuristic. Or the methodology of scientific research programmes, as a fallible means for interpreting science's past, is used by Lakatos to write historical vignettes whose falsifications, or errors, vis-à-vis actual history are provided via Lakatos's ubiquitous footnote apparatus. In addition to formulating a central role for history, Lakatos's pedagogical role is to actively demonstrate the transformation and dissemination of ideas in his work, but with much of that shown rather than said. "Beta," the Teacher commands another character in *Proofs,* "do not constantly heckle Epsilon! Fix your attention on what he is doing and not on how he interprets what he is doing. Go on Epsilon" (116). Let Lakatos go on then too.

The strategy for the remainder of *Guises of Reason* is to explain the Hegelian substratum at work in Lakatos's philosophy from three perspectives: the philosophy of mathematics contained in *Proofs,* the philosophy of science of Lakatos's methodology of scientific research programmes, and the distinctive irrationalism of Stalinist Hungary. Lakatos's philosophy of mathematics and science both build on the Hegelian-Popperian approach to knowledge and criticism as a kind of learning through error—either the errors of a false proof, or the falsifications and contradictions of scientific theories and the research programmes containing them. Here is a brief summary of the explication to follow.

Proofs is Lakatos's philosophical history of nineteenth-century mathematics. The work is organized historiographically like Hegel's *Phenomenology,* as a philosophical-historical bildungsroman, or novel of education and learning. Hegel's book owes much in its form, as observed repeatedly by Lukács, to the narrative genre invented by Hegel's friend Goethe in his *Wilhelm Meister* sketches and books.[30] While not a novel, *Proofs* is similarly organized as a *mathematical* bildungsroman. This historiographical form provides historicizing and pedagogical functions for contemporary mathematics similar to those Hegel intended to provide for Europeans, or at least Germans, and their history up to the early nineteenth century when the *Phenomenology* was published. *Proofs,* more than his philosophy of science, is Lakatos's great philosophical achievement. For millennia, mathematics has been considered the acme of ahistorical, timeless, and usually certain knowledge. Mathematical certainty was decisively lost through the discoveries of modern logic and metamathematics, primarily Gödel's incompleteness theorems. But before Lakatos, there was no systematic account of modern mathematics and its rigor as a fallible form of knowledge built from its own history. Lakatos also does not just propose that as a general thesis but executes his philosophical program with enormous historical detail, rewriting at Cambridge in the late 1950s the backbone of nineteenth-century mathematical history.

Because Lakatos adapts the basic historiographical method Hegel uses to historicize philosophy in the *Phenomenology,* an immediate corollary, so to speak, is that *Proofs* becomes an account of mathematical method in a philosophical tradition entirely separate from the Anglo-American one in which it nominally appeared. It is not that *Proofs* only has "affinities" to post-Hegelian critical theory and historical philosophy; *Proofs* provides the first major bridge between historical philosophy and serious mathematics. It remakes ideas of mathematical method with direct reference to historically fundamental techniques of historical criticism and historicist thinking. At the same time, Lakatos provides an account of how nineteenth-century mathematical progress relates to modern logic and its formalisms as the dominant theory of contemporary mathematical method.

Lukács asserts that the "pre-eminent aim" of Marxist method "is *knowledge of the present*" (H&CC xliii), and while there is nothing especially Marxist in any canonical sense about *Proofs* it is supremely a mathematical present as history. Always historicize, Lukács advised, but nobody expected and few have appreciated Lakatos's application of that to mathematics. Because *Proofs* is written as a series of ongoing proofs, as well as being a history, it also is no reduction of mathematics to history; it is mathematics as history. As well, because Lakatos's account of changes in nineteenth-century mathematics involves a type of mathematical skepticism, and Hegel's historiographical method of the *Phenomenology* is likewise built up via classical skeptical tropes of ancient Pyrrhonism, *Proofs* further revives some of the most influential ideas in the history of philosophy, now reinvented within a historicized philosophy of mathematics. It is Lakatos's virtuoso performance. And because mathematics is not much more than its methods and theorems, being "just" a historical and methodological virtuoso implies that Lakatos is at the center of the subject.

In contrast to the specialized history of *Proofs* Lakatos's philosophy of science is created as a critique of Popper's philosophy of science, analogous to Marx's critique of classical political economy. The critique becomes Lakatos's so-called methodology of scientific research programmes. Critique here means first that Lakatos develops his research programme methodology immanently by critically adopting the concepts, vocabulary, and basic assumptions of Popper's philosophy, just as Marx joined with and worked from within the ideas of an existing and important school of thought. Second, the categories of Popper's and hence Lakatos's methodology are meant themselves to be shown as historical. The philosophy of science, like nineteenth-century political economy for Marx, is characterized by Lakatos as being as much a part of history as science itself. That neither implies that methodological categories are destined to change nor that some autonomous forces or supersubjects of history are creating change. Indeed, the bare conclusion that scientific methodology and its philosophical reconstruction are part of history and subject to change may be something of a banality for the philosophy of science circa 1990 or later. Our interest here is just *how* Lakatos expressed the historical specificity and contingency of Popper's categories and his own. He does so by characterizing the critique of Popper and his own methodology of scientific research programmes reflexively as a historiographical research programme. In so doing, Lakatos ends up creating a conception of philosophical history almost identical to Hegel's philosophical historiography that is, in certain respects, still valid today. Historiography, then, becomes the organizing idea for all of Lakatos's philosophical work, arrived at by a kind of self-application of his own methodology of research programmes to itself. This skeptical trope of self-application—sometimes called self-criticism in political circles—turns Popper's static logic of scientific discovery into a dynamic and changing logic of scientific discovery. Lukács similarly argued in *History and Class Consciousness* that Marxism can and should be applied to itself, thus showing Marxism to be a transient ideology associated with a specific phase of historical consciousness. Combined with his elegant construction of the concept of a research programme via the critique of Popper, Lakatos delivers one fairly complete critical theory of scientific method that even accounts for its own historical becoming. As in *Proofs,* this offers a remarkably creative and instructive account of rationality as product and process.

Within Lakatos's critique of Popper, the key Hegelian ideas he introduces are the omnipresence of contradictions in scientific theories and the role of historical hindsight within scientific criticism. Few descriptions of scientific method give such prominence to contradictions as a positive component of rational process. The failure of formalized models of scientific change are familiar; among other reasons, there is the difficulty of rightly modeling theories and observations that are formally inconsistent. The attitude is often "so much the worse for overly formalized models." But then what happens to Popper's contradictions? They are typically leveled out among the many heterogeneous features of scientific practice. Lakatos's Hegelian perspective is that reasoning through a sea of contradictions is an important feature of scientific reason, meaning exoteric methods of criticism, theory change, or appraisal that are largely independent of who creates, improves, or criticizes the theory or research programme at stake.

Lakatos's positive role for contradictions is wholly antithetical to the received place of formal logic and formalized theories throughout much of the behavioral sciences and theories of educational practice: that contradictions are neither destructive to the growth of knowledge nor decisive for the immediate fate of a theory is far from received cultural wisdom. Nor is Lakatos's notion that scientific method, at least as he describes it, has emerged and evolved over the last half millennium or so. Nor is Lakatos's insistence on the role for retrospective, "backward-looking," historicist interpretations *within* scientific practice and their relevance to our contemporary understanding of major historical episodes. While from a narrow philosophical perspective the bottom-line content of Lakatos's philosophy of science is not as remarkable as it might have been thirty years ago, his systematic organization of contradiction, hindsight, and historiographical method provides a useful corrective to much that is still wrong in received notions of scientific practice. The corrective is a practical one, too, as Lakatos suggests how the methodology of scientific research programmes can be applied to write histories and criticize research programmes. Like the mathematical heuristics of the great Hungarian mathematician and pedagogue George Pólya, or the myriad keyboard exercises *Mikrokosmos* composed by Béla Bartók for his son, the methodology of scientific research programmes is the many-sided toolkit of a world-class educator in scientific method.

And a teacher of historical criticism. Anglo-American philosophy barely has a historicist tradition, much less one that is given theoretical expression in philosophy itself, as happens with Hegel and Marx. "It is common knowledge," notes Lukács again in *History and Class Consciousness,* "that Marx himself conceived this idea of writing a dialectics. 'The true laws of dialectics are already to be found in Hegel, albeit in a mystical form. What is needed is to strip them of that form,' he wrote to Dietzgen" (xlv).[31] But what kind of dialectics? Of social categories, forms of thought, scientific practices, mathematical proofs? Dialectical techniques are as old as the philosophical hills; they occur in many forms, in many philosophers, and are more or less successful in achieving various conceptual goals. What Marx might have been able to provide is also open to question. Lakatos, on the other hand, has given dialectical methods life in the contexts of scientific, mathematical, and philosophical criticism through his critique of Popper and innovative infusion of Hegelian ideas. That whole contribution to the two traditions in which Lakatos's work finds its home really sets his achievement apart, since one of Marx's main accomplishments was similarly to integrate transformed philosophical ideas, many taken from Hegel, within his critique of a specialized discipline: Adam Smith's and David Ricardo's eighteenth- and nineteenth-century, respectively, political economy. This transformative process—itself across multiple traditions and domains of knowledge, now including rigorous mathematics and science—constitutes Lakatos's great expertise. In short, Lakatos provides a general theory of criticism as suggested by Lukács, worked out in detail with applications to important historical episodes from science and mathematics, and addressing Marx's own covert application of Hegelian philosophy to economics.

As far as Marx goes, his role in nineteenth-century political economy has been a perennial problem in the history of ideas and a practical one for the political ideology of Marxism as a science in the once socialist countries. But the many questions of method and criticism answered by Lakatos are completely independent of whether Marx is to be interpreted as a determinist or not, and of the specious metaphysics of dialectical materialism created by Marx's collaborator, Friedrich Engels, and canonized in Marxist dogmatics. At the same time, Lakatos's methodology is created via a critical path starting from Popper, just as Marx's started from Smith and Ricardo. Lakatos does not thereby validate or endorse the content of Marx's economics but he does demonstrate much about the latter's method. The last relevant idea needed to complete this introduction to Lakatos's critical philosophy is obvious once noticed, although Lakatos had only slight overt interest in considering it himself. Popper conceived of his own approach to scientific method as a normative Kantian-positivist effort to demarcate science from nonscience. The search for a normative demarcation criterion is completely preserved in the methodology of scientific research programmes even as Lakatos changes the criteria for what counts as science and reflexively describes conditions for criticizing demarcation criteria themselves. So the obvious question is: Where does Marx fall out as "scientist," taking Lakatos's theory of research programmes as the criterion of science here?

Historically, that question has been answered both positively and negatively, with all manner of justification or dogma, by everyone from Vladimir Lenin to Mao Tse-tung to Louis Althusser. None of them had anything like the sophisticated and well-received version of Popper's critical rationalism created by Lakatos. The answer is that Marxian economics, narrowly yet not artificially defined by that part of Marx's economic theory associated with the labor theory of value and price taken up by Marx from Ricardo and Smith, was conceived and executed by Marx in terms matching key research programme criteria. Marxian economics generated little if anything by way of successful empirical predictions, but that is a criterion for a *progressive* research programme; thus Marxian economics is scientific, but marginally progressive compared to its predecessors. It is important that not any set of ideas, say Thomism or ordinary language philosophy, easily or obviously qualifies as a research programme, so characterizing Marx's economic work in these terms is not vacuous. Moreover, Lakatos's criteria by which Marxian economics turns out to be "science" themselves turn out to be built through historical and critical categories in the same way that Marx built his economic theory. That is all consistent with the received history of economic theory including Smith and Ricardo, with Marx's self-conception of his role vis-à-vis other economists, and everyone's appreciation of the predictive inadequacies of Marxian economics, though even that piece of received wisdom should be considered modulo the relative progress of other nineteenth-century economic theories. The old problem of consolidating the "two Marxes" of political economy and Hegelian-inspired critical social theory is solved through the resources of Lakatos's dual philosophy.

Those, in brief, are some of the lessons of our mysterious Hungarian educator to be explored anon. The next two parts of this book explain Lakatos's Hegelianism, mostly from a sympathetic viewpoint. If individuals or ideas have far richer histories than I could hope to indicate here, such truncated and perhaps potted history is provided in the interest of keeping to the main storyline. I have not pursued several topics related to alienation, totality, mediation, and praxis in Hegel and Marx, while including considerable material going beyond Lakatos's work, principally some history of mathematics, the role of skepticism in Hegel, and the account of Marxian economics just described. The reasons for these choices will become apparent from the conclusions drawn and motivations originating in Lakatos or the relevant history of ideas.

The final chapter, on Hungary, provides a means of understanding just what Lakatos, as a unique cross-cultural educator, was an educator of. The answer is that he was an educator of reason and several of its guises. Some of these guises appear in philosophy—ancient, modern, and contemporary; some occur in the practices of mathematicians and scientists; and some in the political, social, and cultural life of Hungary between World War II and the failed Hungarian Revolution of 1956. The lesson is: learn from them all and beware of their powers. Formulating more precisely just what may be learned from Lakatos about these worlds and their histories is the concluding goal of the last chapter.

"If Faust could have two souls within his breast, then why should not a normal person unite conflicting intellectual trends within himself when he finds himself changing from one class to another in the middle of a world crisis?" wrote Lukács in his reflective and typically self-critical 1967 preface to *History and Class Consciousness.* The change Lukács refers to is his own conversion to Marxism coming from one of Europe's wealthiest families amid Europe's breakdown following World War I and the rise of fascism. Lakatos was also caught between two "conflicting intellectual trends," having left behind the madness of Hungary for the safety and calm of England. Like Lukács, Lakatos was another of the "extraterritorials" of Central Europe displaced politically, geographically, linguistically, and intellectually by the upheavals of the twentieth century, and whose work embodies those changes in an essential way.[32] Lakatos, too, like Lukács staying behind in Budapest, had two souls in his breast. His talent was to know that well, perhaps having implicitly learned it from Lukács, and it is built into his philosophical work. Our goal is not to understand these two souls but the philosophy possessed by them.

1 | A Mathematical Bildungsroman

All this suggests that mathematics and dialectic not only shared a common terminology but were also interconnected disciplines. In fact, it looks as if Plato was writing at a time when mathematics was just a branch of dialectic.—ÁRPÁD SZABÓ, *The Beginnings of Greek Mathematics*

Reason depends on what becomes, the understanding on what has become. The former does not ask: for what? The latter does not ask: from where? Reason delights in development; the understanding wants to arrest everything to use it.—GOETHE, *Maxims and Reflections*

This page intentionally left blank

1 | The Mathematical Present as History

It is remarkable the extent to which the nineteenth century was a time of error for mathematics: not trivial oversights or amateurish confusions but fundamental mistakes in the understanding of mathematical concepts and the formulation of mathematical proofs. These mistakes were not restricted to unknown mathematicians but occurred in the works of great mathematicians such as Joseph Fourier, Augustin Cauchy, and Denis Poisson. Equally notable was the criticism of these errors, which provided the impetus for some of the deepest conceptual reformulations of the century whose influence can be felt down to the present—not only in the specific mathematical fields in which the errors originated but in the foundations of mathematics as well.

In *Proofs,* Lakatos turns this historical observation into philosophical conjectures about mathematical criticism and changes to mathematical method occurring during the nineteenth century. In particular, the philosophy of mathematics represented by *Proofs* demonstrates the historicity of several important features of contemporary methods of mathematical proof. Proof here means informal proof—the proofs of journals, notebooks, and chalkboards—including twentieth-century logic and metamathematics as special cases. What Lakatos calls the method of proofs and refutations expresses different ways in which a theorem and proof can be improved by incorporating attributes of counterexamples or exceptions to a conjectured theorem into the conditions, definitions, or lemmas needed to make a proof meet criteria of rigor. Lakatos's historical thesis is that the method of proofs and refutations became significant to mathematics only after about 1850. His philosophical thesis is that this method is best characterized through an antifoundational and antiformalist philosophy of mathematics based on a historical theory of mathematical criticism, fallibilism, learning, and error. This outlook derives, in part, from Lakatos's adoption of Popper's philosophy of conjectures and refutations from the natural sciences to mathematics. But much of what is philosophically novel in *Proofs* comes from the work's ingenious historiography, the basis of what is entirely Hegelian.[1]

The welter of topics and narratives in *Proofs* tends to amaze and bewilder, so that it is difficult to say at first glance just what kind of philosophical or historical work it is. Overflowing with historical detail, it can appear to have no fixed historical topic. The text consists mainly of intricate historical and methodological analyses of numerous proofs, developed during the nineteenth century, for a relatively simple geometrical result about polyhedra, such as cubes and pyramids, known as Euler's theorem. Yet while Euler's theorem was a prominent theorem of nineteenth-century mathematics, it was not central to the controversies pervading mathematics at that time; other theorems or concepts, such as the representation of functions by Fourier series or the function concept itself, could easily be used to narrate major changes in nineteenth-century mathematics. There is much in *Proofs* on formalized theories and languages that initially is familiar to analytically minded philosophers, yet there is also much philosophical development of mathematical pedagogy and theorem-proving heuristics, topics not normally included in an inventory of problems in the philosophy of mathematics. Lakatos polemicizes at length about the dangers of elitism in mathematics and what may be called epistemological aristocratism; while this polemic is naturally allied with his thinking about mathematical heuristic, without further explanation it would not seem to be directly part of *Proofs*'s thoroughgoing historicism.

Years after *Proofs,* Lakatos's controversial "rational reconstructions," being his historical vignettes of scientific change written strictly from the perspective of his own methodology of scientific research programmes, would become his signature method in the philosophy of science. These historical rewrites would end up driving some readers to exclaim about "Lakatos' absurd historiography" or his "historical parody that makes one's hair stand on end."[2] The history found in *Proofs,* while generally as rigorous as could be demanded by any historian, already uses a similar narrative method, but applied to the history of mathematics rather than the history of science.

When driven by historical scruples back to the philosophy of mathematics, we do not find in *Proofs* any of its traditional problems, such as the definition of number, set-theoretic "existence," the extent of infinite cardinal numbers, or the interpretation of the classical paradoxes. These important though conventional topics provide no starting point here. The solution to understanding the apparent eclecticism of *Proofs*—a combination of history and error, elitism and learning, philosophy and historical parody, and mathematical formalism traced back a hundred years before its time—is to see how a specific historical sensibility is set to work in a characteristically ahistorical subject. As philosophical history, *Proofs* is even described by Lakatos's colleague and arch critic Feyerabend as "the best and most detailed presentation and analysis of conceptual problems in the entire history of ideas," one that "removes the last Aristotelian element, the element of necessity, from modern science."[3] But just how is one to think of mathematical necessity as historically specific and variable? And if history is to be more than novelty or illustration, its use should be tied to some significant mathematical problem or ideas.

The introduction to *Proofs* places it as a challenge to the *complete* identification of mathematics with some formal, metamathematical representation, a view that in some form inhabits nearly every key philosophical position on mathematics since the turn of the twentieth century, and descends from the work of Gottlob Frege, Giuseppe Peano, Bertrand Russell, David Hilbert, Kurt Gödel, and other great logicians. Formalism, in Lakatos's sense, demarcates a broad range of approaches to the philosophy of mathematics, centered on the idea that some largely mathematical theory explains most of what is philosophically important about mathematics: a formalized proof calculus, the output of an abstract computer such as the Turing machine, or a theory of sets interpreted as an "ontology" for numbers, functions, and geometric and other higher-order mathematical objects.[4] Much ordinary mathematical logic, while not necessarily interpreted for its philosophical content, falls under formalism, as one uses specific mathematical conceptions of proof, truth, computation, axiom, and so on to create mathematical theories of mathematics, and these tools have been successfully used to formulate and solve many fundamental problems related to the structure and limits of mathematics. What Lakatos wants to contest is assuming that all methodological problems about mathematics are adequately addressed only by continuing the mathematicization of mathematics. *Proofs* does not reject mathematical formalism but rather recognizes metamathematics and mathematical logic themselves as informal mathematical theories, whose objects of study are not topological spaces, probability measures, or solutions to differential equations but proofs, theorems, formal languages, and their syntax and semantics.[5]

Lakatos's topic is the temporal and historical process of proving theorems and creating new mathematical concepts instead of the formal analysis of existing mathematical theory. Even right in his introduction, Lakatos notes that Gödel's famous incompleteness theorems, to which we will return in chapter 4, are informal mathematical theorems about the limits of formal systems.[6] The heart of Lakatos's challenge is his claim that formalist philosophies of mathematics can say nothing about the historical growth of mathematics or the role of central conceptual changes in many mathematical theories. Formalism, therefore, leaves unexplained changes in patterns of mathematical reasoning, in particular the emergence of nineteenth-century mathematical criticism and its role in creating the mathematical formalisms that are familiar today. An important goal of *Proofs,* then, is to demonstrate the inherently informal character of the main methods of contemporary mathematical proof and their changes through time. As such, *Proofs* is a historical criticism of the limits of modern logical theory in representing its own origins and describing its own rationality. If Lakatos's account is even roughly correct, then by historicizing what has been taken by many to be the standard of a priori or nearly a priori reasoning, he turns centuries of received wisdom about the nature of mathematical knowledge on its head, replacing the traditional quest for certainty or some kind of specialized truth immune to change with an account of mathematical proof itself as a historical phenomenon. Instead of a single, idealized notion of mathematical proof and truth, *Proofs* presents different conceptions of proof changing over time; instead of an explanation of contemporary mathematics in terms of some formalism, *Proofs* provides a historical account of formalism itself.

Others, including W. V. O. Quine, have developed holistic views of mathematics and science in which mathematical logic might *potentially* change through some fundamental reorganization of physics or other empirical study. At one time, similarly, some argued that quantum physics might be better formulated using a "three-valued" logic, allowing statements to be true, false, or indeterminate. But Lakatos is not providing an exotic possibility based on many-valued logic or an abstract holism. He gives a specific history of nineteenth-century mathematics and changes in mathematical practice, following them to the emergence of modern logic—what Lakatos calls the "dominant theory" for contemporary mathematics as a whole.[7] The philosophical source is not a hypothetical empirical world moving us to change our logical theory but the specific historical reality that gave rise to it. Lakatos would likely be open to holistic inferences cast from physics all the way into mathematics or logic. But mathematics is more than a handmaiden to science, and more than a set of theorems and its results. In *Proofs,* mathematics is its results together with their becoming. This historical becoming is just the production of mathematical proofs and the activity of theorem proving in journals, letters, and notebooks, as far as needed to make these part of the intellectual content of mathematics, and leaving out the material, ideological, or other "external" conditions relevant to their causal history. The history of mathematics is taken in *Proofs* not as what mathematics is *about*—say, numbers, structures, or sets, or which particular sets—but as what mathematicians *do,* which is to create, criticize, and improve proofs, create new mathematical concepts, and resolve mathematical conjectures and pose new ones. The history in *Proofs* is neither that of a zeitgeist nor paradigms, nor occult historical forces, nor even people, but history as different ways of creating collections of statements called mathematical proofs. That activity, according to Lakatos, has changed greatly since 1800, and so, in turn, have conceptions of mathematical rigor and truth. Lakatos's ambitious goal is to explicate and enact much of that remarkable historical process, especially the intricate relationships that have developed between mathematical innovation in creating proofs and their modern logical justification: the method of proofs and refutations is supposed to be the engine for that process, in which the competing demands set by mathematical creativity and rigor are traded off to achieve progress. A powerful historical methodology is needed, then, to see mathematics not as the ultimately static and ahistorical subject but as a dynamic historical process. How is this to be attained?

As a case study in the history of mathematics, *Proofs* challenges the one-sided identification of mathematics with its metamathematical representations by showing, as Lakatos says, that "the history of mathematics and the logic of mathematical discovery, i.e., the phylogenesis and the ontogenesis of mathematical thought, cannot be developed without the criticism and ultimate rejection of formalism" (*P&R* 4). So *Proofs* provides a certain kind of history, one associated with "phylogeny" and "ontogeny," Ernst Haeckel's correlative terms for biological development in species and the life of an individual, respectively. Lakatos cites Henri Poincaré and George Pólya as mathematicians who have proposed that Haeckel's now-discredited "biogenetic law" be adapted for mathematical pedagogy. The motto of the biogenetic law is that "ontogeny recapitulates phylogeny," meaning that the biological growth of the individual repeats in abbreviated and caricatured form the historical development of the species, thus telescoping millennia of historical change into the physical development of a single life. As a *narrative* technique, and not a historical or even psychological "law," the approach means that the "embryonic" development of a student's knowledge should recapitulate in brief the history of mathematics itself, or at least those parts relevant to the mathematics being taught. The biogenetic law is not mentioned again in *Proofs* but it defines the architectonic of the whole work, and Lakatos's single remark characterizes the book's fantastic historiography.

The entity whose "ontogeny" is followed in *Proofs* is nothing mental or mindlike, but rather a fully externalized discursive object: it is simply Euler's theorem itself, stating that the number of vertices plus the number of faces of a polyhedron exceeds the number of edges by two: $V - E + F = 2$. The "phylogeny" against which this individual history is laid out is mostly that of nineteenth-century theorems and their proofs—many explicitly falsified, some not—of Euler's classic result. Lakatos develops this history using two techniques executed with great virtuosity: a dialogue form in which Euler's theorem is continually debated and reproven; and a running commentary in the footnotes associating the dialogue's timeless and analytic content with events mostly from nineteenth-century mathematics, but including contemporary and ancient sources as well. Both of these narrative methods combine to give *Proofs* its specific philosophical-historical form.

The eighteen characters of Lakatos's dialogue—named Alpha, Beta, Gamma, and so forth—debate the status of Euler's theorem from its birth in the seventeenth and eighteenth centuries as a naive conjecture proposed by René Descartes and Euler, to its deep formulation in algebraic topology created by Henri Poincaré in 1899. The character Alpha is a constant creator of counterexamples to Euler's theorem. Beta and Sigma represent methodological aspects of both Cauchy's and Niels Abel's mathematics, a method Lakatos calls "exception-barring" that plays, he argues, a key role in the mathematics of the early nineteenth century. Delta is a "monster-barrer" who repeatedly contracts and redefines definitions of polyhedra when faced with Alpha's counterexamples. Gamma is a naive falsificationist who gives up his conjectures as quickly as Alpha appears to refute them. Lambda, like Beta, is one of the most important shapes of mathematical criticism—Lambda introduces the methodological pattern called "lemma-incorporation" and represents, for Lakatos, the major change in mid-nineteenth-century mathematics. Iota is an impish dialectician with only two lines in the entire dialogue. Epsilon, who delivers Henri Poincaré's modern proof, is a rigorous formalist who claims to provide absolute certainty for his almost purely syntactic version of Euler's theorem—Epsilon in many respects is *us* and provides the greatest challenge to the role given by Lakatos to history and informal mathematics. Rho offers psychologistic arguments reminiscent of L. E. J. Brouwer's mathematical intuitionism. Pi is a philosopher of language who adroitly enters late in the dialogue to reinterpret its beginnings. Kappa and Theta uncover radical skepticism within mathematical proofs. Finally, along with the Teacher, this cast lives out a century of mathematical history as they make their contributions to Euler's theorem, reformulate their reformulations, and themselves undergo the metamorphoses brought about by the revolutions in rigor of Cauchy, Karl Weierstrass, and other nineteenth-century mathematicians.

What is notable about all these characters, save perhaps for the Teacher, is that each one is a truncated, condensed, and often ironically caricatured version of historically paradigmatic mathematical and philosophical standards. Each such pattern of mathematical reason is wedded to specific historical events through Lakatos's frequent and expert use of direct quotation from historical documents, sometimes slightly yet crucially modified when ventriloquized through one of the rapidly changing voices, but then "corrected" in the footnotes below. In his introduction, Lakatos tells us that these shapes of mathematical heuristic and criticism are not to be confused with actual history: the dialogue "is meant to contain a sort of rationally reconstructed or 'distilled' history" (*P&R* 5). So, for instance, while "Alpha, Beta, and Gamma suspected three lemmas [that is, supporting theorems] when no global counterexamples turned up, . . . in actual history proof-analysis came many decades later: for a long period the counterexamples were either hushed up or exorcised as monsters, or listed as exceptions" (*P&R* 48n1). Or, in another context, "Pi's statement, although heuristically correct (i.e., true in a rational history of mathematics), is historically false. (This should not worry us; actual history is frequently a caricature of its rational reconstructions)" (*P&R* 84n2). Such historical and metahistorical remarks occur in Lakatos's elaborate footnote apparatus, not in the dialogue proper, and their role is to mediate between the complicated reconstruction of the text and the actual history of mathematics: "The real history will chime in the footnotes, most of which are to be taken, therefore, as an organic part of the essay" (*P&R* 5).

One may end up disagreeing with his history of nineteenth-century mathematics and skeptical claims about its strange consequences for twentieth-century formalism. But Lakatos's differential historiography, while not completely original, makes *Proofs* one of the earliest, most radical, and ingenious twentieth-century philosophical texts to actively exhibit the importance of narrative choice and theory ladenness for historical inquiry.[8] Another way of describing Lakatos's biogenetic technique and its unusual historiography is to recognize that *Proofs* is composed as a mathematical and historical bildungsroman, a "novel of education," the "hero" of which is Euler's theorem. The simple reformulation, justified even historically, will allow us to track Haeckel's biogenetic law, as Lakatos's shorthand for his technique, back to the philosophical concept of *Bildung,* its expression in the first bildungsroman, and then to Hegel's similar use of the narrative technique and its underlying philosophy of error in the great bildungsroman of Western philosophy, his *Phenomenology.* The key to *Proofs* is that Lakatos does not just flirt with the occasional Hegelian "mode of expression," like the mention of "ontogeny" and "phylogeny," but that a historiographical method of central importance and in constant use stems directly from Hegel's *Phenomenology.*[9] This implicit appropriation of Hegel, intentional or not, puts *Proofs,* as philosophical history, decades ahead of its time in philosophical sophistication and is part of Lakatos's great contribution as a philosophical exile.

We will return to the explicit challenge set by Lakatos for mathematical formalism in chapter 4. The task now is to reconstruct *Proofs* starting with an account of its historiographical form. Following this, the conceptual and historical details of the method of proofs of refutations will be examined, as well as the role of skepticism in it. The Pyrrhonian skepticism at work in *Proofs* turns out also to be the engine of historical criticism found in the *Phenomenology.* Both the overall historiographical style of *Proofs* and the principles of the method of proofs and refutations, then, are based on much the same philosophical foundation through which Hegel provides his constructive historicist account of Western philosophy, society, and culture. Gödel's second incompleteness theorem is presented in chapter 4 as a concrete mathematical example of Lakatos's mathematical skepticism, while also illustrating Lakatos's account of the translation process between informal and formal mathematics. This chapter closes with a return to the preface to the *Phenomenology* and Hegel's portrayal there of the book's antiromantic pedagogical purposes; these goals largely reappear in Lakatos's criticism of a Euclidean-deductive style in the presentation of mathematical proofs, to which we return in chapter 5.

As Lakatos observes in opening *Proofs,* "It frequently happens in the history of thought that when a powerful new method emerges, the study of those problems which can be dealt with by the new method advances rapidly and attracts the limelight, while the rest tends to be ignored or even forgotten, its study despised. This situation seems to have arisen in our century in the Philosophy of Mathematics as a result of the dynamic development of metamathematics" (*P&R* 5). Like much else that Lakatos wrote, that beginning sentence is rife with the meanings of history.

Hegel's Philosophical Bildungsroman

Engels initiated the use of Haeckel's law in the Marxist tradition as a description of Hegel's architectonic in the *Phenomenology.* In his *Ludwig Feuerbach and the End of Classical German Philosophy,* Engels says of the

Phenomenology that "one may call it a parallel of the embryology and the paleontology of the mind, a development of individual consciousness through its different stages, set in the form of an abbreviated reproduction of the stages through which the consciousness of man has passed in the course of history."[10] Engels's scientistic reformulation of Hegel's technique reproduces two aspects of Bildung as found in German romanticism and science at the end of the eighteenth century: the idea of abbreviated or condensed forms of the past appearing in the present, and the integrated appearance of those condensed forms in contemporary individuals. Although not using Haeckel's terminology of one-half century later, the anatomy teacher Karl Kielmeyer—a friend of the zoologist and paleontologist Georges Cuvier, who correlated anatomical and paleontological forms—appears to have formulated Haeckel's principle as early as 1793.[11] At the same time, the biogenetic principle was regarded by romantic writers and philosophers as an important new scientific principle, and was paralleled by new forms of universal history and the emergence of the bildungsroman as the law's literary analogue. Both the biological and literary interpretations of the biogenetic principle were used by Goethe, who was a natural scientist as well as a great writer, and suggest how the biogenetic law could serve as the basis for a new literary form. Goethe's *Urpflanze,* for example, was a speculative archetype of all plants, representing plant life in general, and the idea of such a synecdochic object, a unique entity with universal characteristics, would become one of the most powerful, and also misconceived, notions of nineteenth-century natural science. But this theoretical understanding of natural morphology and metamorphosis also provided a legitimate narrative technique through which a general concept of human character could be represented by a single typical individual, notwithstanding his or her specific historical attributes—such as the hero of the first bildungsroman, Goethe's 1795 *Wilhelm Meister's Lehrjahre.*

The bildungsroman provides a model for following the development of individual character while simultaneously painting a portrait of the culture from which a hero gains values, skills, and interests. It takes us along an educational path beginning with a naive, *ungebildet* hero, and ends when this hero has reached a certain level of maturity attained through several cumulative and character-forming experiences. The main concern of Goethe's novel is Wilhelm's education, which Goethe develops by constructing a milieu of figures forming, as Schiller calls it, a kind of "beautiful planetary system," in which characters appear only insofar as they are needed to help portray Wilhelm's growth from inexperience to maturity.[12] Wilhelm's educational process is directed by his experience with error, as described by the Abbé, a philosopher of pedagogy: "Not to keep from error is the duty of the educator of men, but to guide the erring one, even to let him swill his error out of full cups—that is the wisdom of teachers. Whoever merely tastes of his error, will keep house with it for a long time and be glad of it as of a rare good fortune; but whoever drains it completely will have to get to know it, unless he be insane."[13] Wilhelm's grasp of reality comes about only through the changes and possibilities afforded him in the theater, and by allowing his emotional and intellectual life to go astray. These errors correspond less to Goethe's idea of natural metamorphosis, by which error is analogized with bodily disorders to be overcome by health, than they do to Goethe's conception of the role of error for those "outstanding men [who] have been able to extend the sciences . . . if we make out the ways in which they have been sidetracked and in which a large number of disciples, sometimes for centuries on end, have followed them until later experience put the observer on the right path again."[14] Error, therefore, is just the basis for temporal learning and progress, and is neither a biological nor metaphysical principle.

Wilhelm is also a typical young man, not a genius, and his role combines characteristics common to all young men and yet, in the novel, that depend on Wilhelm's circumstances only. "It is true symbolism," commented Goethe elsewhere, "where the specific represents the general, not as a dream and shadow, but as a living instantaneous revelation of the unfathomable."[15] Wilhelm Meister, while not plumbing the depths of the "unfathomable" as would Goethe's *Faust,* is nonetheless such an instance of the specific representing the general, for Goethe's purpose is not to describe a particular case of pedagogy but a typical one in its relation to the hero's social context, primarily the German theater of Goethe's time. Wilhelm's progress through several love affairs and the possibilities, or lack of them, to succeed as an actor, culminate in his maturity. It is that self-formative process bounded by a changing social context, and Wilhelm's eventual maturity, that make up his Bildung: simultaneously a result and the process that created it. Wilhelm's Bildung combines his self-cultivation, his development through interacting with his environment, and the final "shape" he assumes at the end as an ordinary citizen.

The first novel of education, therefore, contains the primary narrative devices for explaining formative growth predicated on a process of learning through error: the representative though indistinctive hero; the stock characters used to portray the hero's milieu, but not in depth; where needed by the author, a reflective voice like that of the Abbé to summarize a particular vision; and the intertwining set of personal experiences of individual and social failure, growth, and change, often described with Goethe's irony to show a reader what Wilhelm will only come to learn later—the irony being used to phase the reader's understanding and learning in parallel to the Bildung described in the novel. As we shall see, several of the central narrative devices that appear in great abstraction in the *Phenomenology,* and then again in *Proofs,* appear in *Wilhelm Meister,* the primary differences in the philosophical works being a broader historical perspective and a more abstract conception of progress through error.

The step back to Goethe and *Wilhelm Meister* thus highlights in more prosaic terms a common formal structure used to explain education in time, and importantly, how the end result is related to the *Bildungsprozess* itself. Should we understand Wilhelm to be "guided" teleologically to the goal, namely, the self he has come to be at the novel's end, so that his life has unfolded like a flowering plant in Goethe's garden? No. Wilhelm is goal oriented like all humans: he wants to act in the theater, to experience adventure and love. His learning occurs through the mismatch between his expectations and reality, but as in our modern understanding of natural selection and evolution, there need not be any overall force or telos guiding the entire process. Wilhelm has advanced toward *a* goal, as he and we can see in retrospect, even though the goal was nowhere there during his education—certainly not for him, and only perhaps implicitly in his environment. His educational path consists in the progress of his education, more than its effect per se, but then he is what he becomes. Wilhelm's apprenticeship demands mastery of skills and his world, but as Schiller maintained,

> This idea, which can only be the result of mature and fulfilled experience, does not itself guide the hero of the novel; it cannot and must not stand before him as his end and goal; for if he were to think of the goal, he would have *eo ipso* [by the very fact] attained it: it must stand as a signpost *behind* him. In this way, the whole acquires a beautiful purposiveness without the hero having a purpose; reason finds a task fulfilled while imagination completely maintains its freedom.[16]

This freedom is perhaps not *complete,* as the social milieu is the real source of "refutations" to Wilhelm's "conjectures," and Goethe's message is that the hero cannot become anything he wishes.[17] Nonetheless, Schiller aptly described the balance in Bildung between purposiveness, retrospective purpose, and responsiveness to error, much as is needed to understand the role of teleology for Lakatos's history of the mathematical present.

In comparing Hegel and Popper's skeptical starting points for their philosophies in the introduction, we saw that the pedagogical philosophy represented by the Abbé's conception of learning through error in *Wilhelm Meister* is also the start for Hegel's *Phenomenology,* which Hegel depicted as a "detailed history of the *education* of consciousness itself to the standpoint of Science" (*PhS* 50)—that is, a philosophical history of forms of Bildung. In Hegel's historical introduction to Western philosophy, in which the reader is expected to work through numerous stylized theories of knowledge as well as forms of societal organization and self-knowledge, the fundamental idea is that human consciousness is inherently historical: its contemporary form is the result of its own lengthy Bildung, and *that* philosophical perspective on human being should be taught by reorganizing the principal modalities of relevant individual and social knowledge in a pedagogically useful historical sequence. The "hero" of the *Phenomenology* is a generic, ordinary consciousness—you or me, or early nineteenth-century analogues—portrayed through so many stylized ways of knowing nature, the social world, the human body, history, all as forms of self-knowledge, and where the "self" may be an individual knowing subject, a community, or a political entity. These *Gestalten,* or "shapes" of knowledge as Hegel called them, all of which are taken from the history of philosophy and the West—from the pre-Socratics and Greek polis through the Enlightenment to Hegel's present—are, like the characters of *Wilhelm Meister,* representative of their type, historically shallow but real nonetheless. They are presented as caricatures to highlight what for Hegel was characteristic, even critical, of their way of knowing the world and their shortcomings or generalized errors. We, the readers of the *Phenomenology,* "observe" these errors, as Hegel said, and Hegel's reconstruction of the sequence of patterns of knowledge or consciousness can be construed as the Bildung of Western philosophy and society to Hegel's philosophical present. The series of types or gestalts of knowledge—from Hegel's Sense-Certainty to Perception and Kantian Understanding, through varieties of individual and social self-consciousness and social organization, through the broadly reflective representations of art, religion, and philosophy—are thus integrated, like so many of Wilhelm's adventures and experiences, into Hegel's description of the shape of contemporary philosophical knowledge, and guidance for an individual understanding of the roles of different types of knowledge in the present; and these shapes of knowledge still are intended to represent actual history as well.[18]

The *Phenomenology* is intended simultaneously as a history of Western philosophy and European society, and a characterization of the principal structures of modern European consciousness, all the result, up to Hegel's time, of a many-layered Bildungsprozess. The work is a historical itinerary of self-conceptions, or an inventory of what one should know to think about problems of the self, society, and culture. The educational process like that Schiller portrayed for *Wilhelm Meister,* ends with *a* goal, but that goal is only one we have defined from our standpoint in the present: no demiurge of history is assumed here. Even the sequential changes or transitions of one pattern of consciousness to another, say from Hegel's caricatured Stoicism to Skepticism, and then to the Unhappy Consciousness, is, though taken from history, *our* reconstruction of it: "This way of looking at the matter is something contributed by *us,* by means of which the succession of experiences through which consciousness passes is raised into a scientific progression—but it is not known to the consciousness that we are observing" (*PhS* 55).

Thus from the simplicity of Goethe's generic hero, Wilhelm Meister, the *Phenomenology* abstracts up to a generic yet ordinary human personal or social consciousness—really any student of philosophy, culture, or history, or a "typical" German—and then back again with a similar purpose though smaller scope in *Proofs* to a representative theorem as hero, Euler's theorem. Not a "genius" theorem, like the representation of functions by Fourier series, but one comprehensible by high school students, and providing an exoteric pedagogy of mathematical method that reflects the real history of nineteenth-century mathematics. The shapes of mathematical consciousness or forms of knowledge in *Proofs* are similarly caricatured. They are not intended as histories by themselves but as representing the education of nineteenth-century mathematics and new instances of Schiller's "beautiful planetary system": for example, "monster-barring" and "monster-adjustment," as Lakatos calls two of the simplest patterns for dealing with counterexamples to mathematical conjectures; a somewhat more sophisticated "exception-barring"; then the method of proof and refutations, or what is also termed "lemma-incorporation"; then a more sophisticated proofs and refutations; unnamed variants of mathematical skepticism; and the formalist Euclidean pattern typical of contemporary mathematics. This reconstructed series, which we will explore shortly, is what the student-reader learns by working through the proofs of Euler's theorem, some of them quite challenging, just as Hegel asked his readers to experience the "exertion" of philosophical concepts. The *Phenomenology* is philosophy for its readers, just as *Proofs* is mathematics; reading each requires an apprenticeship in what philosophy or mathematics is, by demanding that the reader engage in their methods. Both texts are, as Hegel claimed in the preface to the *Phenomenology,* a "ladder" and "already Science" (*PF* 40, 56) through which the reader's understanding of the history of philosophical or mathematical consciousness becomes a caricatured instance of the process of which he or she reads. At the same time, and going beyond the limited role of the theater and German culture in *Wilhelm Meister,* the cumulative narrative of *Proofs* is one of contemporary mathematical rigor, just as Hegel's history is one of modern consciousness. The self-formative, constructive, or constitutive role for historical experience in mathematical method, therefore, is implicit in the historiography of *Proofs* as well. Mathematical rationality, like rationality for Hegel, lay in the process creating a result, not just the result, and this process is driven by progress through error. As Lakatos contends, he wants to correct philosophers and "modern rationalist historians of mathematics" who treat the growth of knowledge "on the basis of unchanging methodology" (*P&R* 139n1, 134).

How, then, should one represent a changing methodology showing mathematical progress through error, which is historically based, describes the present or part of it, and is pedagogically useful? Hegel's detailed answer is the biogenetic narrative as an abstract and generalized bildungsroman of consciousness.

"The task of leading the individual from his uneducated standpoint to knowledge had to be taken in its general sense, and the general individual, the self-conscious spirit, had to be considered in its education" (*PF* 44), by which Hegel meant that the reader's pedagogy is provided via a history of the education of contemporary consciousness in general. Hegel went on to explain:

> As for the relation of the two: in the general individual every moment shows itself as it gains concrete form and its own shape. The particular individual is the incomplete spirit, a concrete form in whose whole existence one determination predominates, while the others are present only in blurred features. In the spirit who stands on a higher level than another, the lower concrete existence has been reduced to an insignificant moment; what formerly was the matter itself has become a mere trace; its form is shrouded and become a simple shade. (ibid.)

That is, as historical knowledge progresses, earlier forms of knowledge are reduced to "moments," constituent parts whose inadequacies are left behind, as are actual historical details, creating a kind of caricature. The student or reader then recapitulates the past:

> Through this past the individual whose substance is the spirit that stands on a higher level [that is, contemporary culture] passes in the

> same manner in which the student of a higher science goes once more through the preparatory knowledge that he has long mastered, to present the contents to his mind: he recalls these memories without being interested in them for their own sake or wishing to abide in them. The individual must also pass through the contents of the educational stages of the general spirit, but as forms that have long been outgrown by the spirit, as stages of a way that has been prepared and evened for him. (ibid.)

The history is necessarily truncated: "Thus we see that as far as information is concerned, what in former ages occupied the mature spirits of men has been reduced to information, exercises, and even games suitable for boyhood; and in the boy's pedagogical process we recognize the history of the education of the world as if it had been traced in a silhouette" (ibid.).

These historical silhouettes or caricatures are Hegel's gestalts of knowledge, or Lakatos's mathematical characters and heuristics, and also the basis for contemporary learning and the continuation of the cultural present: "This past existence is property that has already been acquired by the general spirit which constitutes the substance of the individual and, by thus appearing to him externally, his organic nature" (ibid.). Hence education, "considered from the point of view of the individual, consists in his acquiring what is thus given to him; he must digest his inorganic nature [that is, his given social and historical world] and take possession of it himself. But from the point of view of the general spirit as the substance this means nothing else than that this should acquire self-consciousness and produce its becoming and reflection in itself" (*PF* 46). The student learns by internalizing a representation of the history that created the "general spirit," the contemporary milieu in which one finds oneself. If such a historical reconstruction using the biogenetic metaphor is one approach to attaining historical consciousness and demonstrating the historicity of consciousness, then Lakatos has achieved that goal for mathematics.[19]

Set against these simultaneous pedagogical and historical goals, the complex and unorthodox histories of both the *Phenomenology* and *Proofs* are understandable, as are exasperated responses, such as Rudolf Haym's reaction that

> the *Phenomenology* is psychology reduced to confusion and disorder by history, and history deranged by psychology. . . . History in the *Phenomenology* is a history emancipated from the rule of chronology. . . . [F]ormations remote from each other are suddenly juxtaposed in obedience to [Hegel's] order, while others which condition each other historically are

> wrenched apart. One is equally baffled whichever strand one tries to hold on to.[20]

Hegel had indeed been practicing for the approach taken in the *Phenomenology* for many years, including a rewriting of the Gospels in terms of Kantian moral psychology from his time as a theology student. As with Lakatos, history had to be intentionally falsified for pedagogical and philosophical purposes. H. S. Harris notes that in "Hegel's mind this did not involve any falsification of the historical record, and we can safely assume that he is not consciously or deliberately guilty of falsification. Rather he is seeking to undo a certain 'falsification of the record' which is inherent in its existence as a 'historical record' at all. His account is not meant to take the place of the historical record, or even to be read independently of it, but rather to throw light on it."[21] The rewriting of history, along with its partial falsification needed for *Bildungsphilosophie,* and the criticism of history guided by philosophically untenable assumptions are thus implicated from the start in this historical philosophy. Lakatos, for example, also points out repeatedly how histories of mathematics falsify errors in the history of Euler's theorem, which Lakatos attributes to the implicit assumption that mathematics is infallible and does not progress through error.[22] Lakatos is far more explicit about this process than Hegel, using the footnote apparatus to detail actual history, criticize other historians, and comment on the dialogue.

But in *Proofs,* Lakatos still only practices his rewriting of history. His criticism of falsified history as well as a historiographical account of the roles of historical caricature and interpretation will have to wait for over a decade and the methodology of scientific research programmes.[23] The history of either *Proofs* or the *Phenomenology,* in which historical events may be rearranged in the dialogue "before" their true chronological order, also differs from the more orderly world of *Wilhelm Meister* and marks an important philosophical change. Continuing the literary parallels here, the narrative strategy of the *Phenomenology* is a cross of structural elements from both *Faust* and *Wilhelm Meister.* While appropriating the pedagogical model of the latter to emplot the development of a generic philosophical consciousness, Wilhelm's ordinary surroundings of the German theater and his evolving love life are replaced with a different kind of social and historical backdrop, to which Rudolf Haym is reacting and within which Hegel's education of *Geist* is carried out. This backdrop is of the kind found in *Faust:* an abbreviated, highly condensed, and truly fantastic reconstruction of human history through which Faust, erringly led along by Mephistopheles, makes the journey recapturing the classical world as a means of revitalizing the present.

As a "drama of the human species," Faust's eternal striving represents the destiny of Western humanity, but the figures appearing in the poem showing this are one-sided and generic, combined, as Lukács says, in "a fantastic discontinuous, subjective-objective time and time-sequence just as in the *Phenomenology*."[24]

This description applies to *Proofs* as well, with its three overlapping sequences of historical time: the reconstructed time of the dialogue, in which actual history is radically foreshortened and permuted to express the abstract methodological content of various proofs of Euler's theorem; the time of the actual history "chiming in the footnotes" (*P&R* 5); and finally, the reader and narrator's present time in which both these histories are presented, Euler's theorem is given as a mathematical proof, and the entire story is made into an account of the mathematical and philosophical present. In its passage from an ungebildet, naive thought experiment to a modern theorem, the journey taken by a mathematical conjecture reveals its accommodation to contemporary criteria of rigor, themselves being the end product of a complex historical evolution. Euler's theorem, as the "generic" hero of *Proofs*, represents both a definite historical time and the general constructive relationship of an individual theorem with its mathematical history. The dialogue and footnote form of *Proofs*, like the world of the theater in *Wilhelm Meister*, yet closer to the multifarious spirits in *Faust*, through its constant movement between contemporary wisdom and the historical past, represents the mathematical world of informal theorems and proofs as a heterogeneous complex in which various historical stages, levels of rigor, and competing notions of theoretical sophistication exist unevenly side by side and mutually interact. As Goethe put it in *Faust:* "Consider it and you will better know / In many-hued reflection we have life."[25]

Against Epistemological Aristocracy

The historical and conceptual background of error, learning, Bildung, and historiography needed to explore *Proofs* is now complete. Goethe and Hegel's shared philosophy of error is not a vague weltanschauung but the foundation for the pedagogically motivated historiography of the *Phenomenology*, summarized as the "biogenetic" technique also found in *Proofs.* The introduction to the *Phenomenology* provides the details of Hegel's so-called phenomenological method, in which Hegel effectively defines his historiographical method as a mitigated skepticism. Hegel neither rejects skeptical antifoundationalism nor affirms any dogmatic epistemology as an apodictic basis for any kind of knowledge. We will return in chapter 3 to Hegel's phenomenological method and its skeptical fallibilism in a discussion of Lakatos's method of proofs and refutations. There is, though, another motivation for Hegel's *Phenomenology* needed for Lakatos's philosophy of mathematics and science: Hegel's framing of the *Phenomenology*'s pedagogical goals through his polemic with his contemporary and onetime collaborator Schelling.

Hegel's twofold aim in the *Phenomenology* of offering practical pedagogy within a history of consciousness is repeated in Lakatos's theory of mathematical heuristic as it emerges from the history told of Euler's theorem on polyhedra. Where Hegel provides a philosophical pedagogy, Lakatos delivers a theory of learnable proof techniques.[26] To accomplish this, the narratives of their books become embodiments of the history they retell. For Hegel—and as we shall see, in Lakatos's criticism of the Euclidean style of presenting proofs—this strategy is intended to be exoteric, in principle accessible to any student willing to expend the effort. Rationality is marked by the individual's internalization of historical content implicit in the contemporary world and the historical process through which the knowledge learned was created. The individual—certainly any reader of the *Phenomenology* for Hegel—has, he asserted, "the right to demand that science should at least furnish him with the ladder to this standpoint—and show him this standpoint within himself" (*PF* 40). The *Phenomenology* and *Proofs* are very much about bootstrapping the student into learning how to learn. This approach is in stark opposition to what Hegel saw in Schelling as an antipedagogical, anticonceptual, and elitist conception of knowledge. "Others," Hegel wrote, meaning the unnamed Schelling,

> say that truth exists only in that, or rather as that, which is called now intuition, now immediate knowledge of the absolute, religion, or being. . . . It follows that what is then demanded for the presentation of philosophy is the opposite of the form of the Concept. The absolute is supposed to be not comprehended but felt and intuited; it is not its Concept that is meant to prevail and be proclaimed but its feeling and intuition. (*PF* 14)

Intuition, inarticulate subjectivity, and the nondiscursive spark of genius are forms for Hegel, literally, of unreason because they are without process, outside of time, and not obviously communicable across generations or among the members of a culture. The idea of a philosophical intuition is anathema for Hegel, pitting an undifferentiated and indescribable skill or tacit knowledge against the concrete "determinations" or varied patterns of historical consciousness.

"In the absolute," or Schelling's "A = A," Hegel explained, "everything is one," and "to pit this one piece of information against all the distinctions of knowledge, both attained knowledge and the search and demand for knowledge . . . that is the naïveté of the emptiness of knowledge" (*PF* 26). Without the elaborate exposition given by Hegel, philosophy "has the appearance of being an esoteric possession of a few individuals" (*PF* 22), and cannot ever hope to provide the constructive, self-formative account of human reason, reflective capacities, and institutions that is Hegel's goal. Schelling, at least for a time, had believed that philosophy is necessarily esoteric, having taken up a hypothetical idea of Kant's in his *Critique of Judgment* of a cognitive faculty that would overcome the inescapable difference between the particulars of phenomenal experience and universal representations of conceptual understanding.[27] Kant called this hypothetical means of knowledge an *intellectus archetypus,* which Schelling effectively affirmed as an empirical reality, perhaps influenced by the popular romantic cult of genius or mystical ideas more generally.

The *Phenomenology* and its historical logic was intended as an antidote to Schelling's implicit cult of genius, talent, or personality. Schelling's intellectual intuition is immediate, meaning it is not mediated through education or history. He allowed that art and religion were legitimate means of philosophical understanding, but from Hegel's perspective, outside of these Schelling created a clear breach of communication between philosophy and all else. As a kind of philosophical knowledge given to those fortunate to be born with it, it is, for Hegel, entirely unrelated to ordinary modes of knowing, a special kind of cognition altogether. Consequently, certain types of knowledge are effectively unlearnable. Another result is that the world becomes divided between those who have the inarticulate feeling or intuition to grasp the absolute and those who inexplicably lack that capacity. Schelling, as Lukács remarks, created in Hegel's eyes an "aristocratic theory of knowledge" in which authentic knowledge may be accessible only by a chosen few. In his lectures, Hegel made this aristocratism clear:

> As the science of reason, philosophy is of its very nature . . . available for all. Not all can attain to it, but that is not in question, any more than that few people become princes. That some men are placed higher than others is a scandal only if it is claimed that they are creatures of another kind and that nature had created them so.[28]

Among the reasons for the form of Hegel's *Phenomenology* is his intention to provide a concrete substitute for the irrational impulses implicit in Schelling's view. The pedagogical themes of the *Phenomenology,* including Hegel's "negative" learning through error and the biogenetic narrative style, were already "in the air" in various works of Goethe, Schiller, Lessing, and Fichte, but the strongest single expression of why these combined ideas are necessary for philosophy is found in Hegel's attack on Schelling's elitism. Without learnability, the entire premise of Hegel's historical philosophy breaks down and Bildung makes no sense. One cannot believe in a conception of historical self-formation, of both individual and social subjects, when there is no articulated or at least communicable process for passing knowledge from one generation to the next. If one conceives of rationality as a historical process, as well as learning shared among individuals and the societies they make up and reproduce, then Schelling, or at least the unnamed stand-in found in Hegel's preface, has no learning, language, society, rationality, and self-consciousness at all: Schelling's irrationalism denies the worldview based on Bildung and actively works against it. What matters for Hegel is history and its detailed exegesis to correct the mistaken belief that philosophical knowledge is a gift or privilege:

> Without this elaboration [*Ausbildung*] science lacks universal intelligibility and has the appearance of being an esoteric possession of a few individuals. An esoteric possession: for it is present only in its Concept—only its inside is there. Of a few individuals: for its inarticulate appearance makes its existence merely individual. Only what is completely determinate is at the same time exoteric, comprehensible, and capable of being learned and of thus becoming the property of all. The intelligible [*verständlich*] form of science is the way to science which is offered to all and made equal for all; and to reach rational knowledge by means of the understanding is the just demand of consciousness as it approaches science. . . . Above all, it [Hegel's approach] will differ from that enthusiasm which, as shot from a pistol, begins immediately with absolute knowledge, having done with other standpoints simply by declaring that it will not deign to take notice of them. (*PF* 22, 42)

Versions of intellectual intuition, or conceptions of knowledge implying the need for specialized and inarticulate abilities, or antipathy toward discursive practices generally in favor of tacit understandings, or forms of knowledge without apparent means for realizing or maintaining themselves as historical processes: these are all possibilities that are anathema to someone like Hegel or Lakatos who understand reason as an articulated process and historical growth. Schelling would call his later philosophy "existential," and by virtue of being addressed in Hegel's preface, it is a contemptible, not even "scientific" philosophy. Hegel curiously began his preface by commenting on the inappropriateness of having prefaces in a rigorous philosophical work at all: part of this polemic means that Schelling's inarticulate intuitionism is to be cast out, as it were, excluded from the philosophical city like a scapegoat, just because he or she does not know how to live by human reason in a polis.

Lakatos, we shall see in chapter 5, erupts in polemic and argument against such intimations both in his philosophy of mathematics and science, just as laid out by Hegel's polemic at the start of his philosophical pedagogy. In contrast to Hegel against Schelling, Lakatos's contentions in the philosophy of mathematics involve ideas far deeper and more influential than Schelling's "A = A," comprising an attack on no less than the standard method of Euclidean-deductivist-style definition-theorem-proof as a canonical means for communicating and explaining mathematical knowledge. The Euclidean style and its formalist ideology, from this perspective, is a species of irrationalism in the guise of reason. That remarkable critique in Lakatos would be misunderstood if taken as mere polemic accompanying his implicit appropriation of Hegel's Bildungsphilosophie. Lakatos's repeat performance of Hegel's attack on Schelling is thoroughly grounded in his method of proofs and refutations, the subject of the next chapter.[29]

This page intentionally left blank

2 | The Method of Proofs and Refutations

Mathematical *theorems* as they appear in history for Lakatos are fallible *conjectures* about some domain of mathematical entities or objects, the latter terms used, as in ordinary mathematics, as a way of speaking about polyhedra, numbers, differential equations, logical theories, stochastic processes, sets, or vector spaces as concepts without any special ontological import. An informal conjecture, on the other hand, like Euler's theorem stating that $V - E + F = 2$ for polyhedra, or any other ordinary mathematical conjecture or theorem, is quite real and exists as a part of history as much as the Declaration of Independence. Like a scientific conjecture, the theorem may be erroneous or false, even after a proof is created and published, and much of *Proofs* is taken up with showing how numerous false versions of Euler's theorem were promulgated during the nineteenth century. Pedagogically, even simple or known theorems, for which uncontested proofs have long existed, are conjectures for the student tasked with proving them as exercises. Mathematical fallibilism in *Proofs* means the general view that all informal mathematical theorems are potentially falsifiable conjectures, before or after their proofs are discovered and accepted. Neither for the student nor the mathematical community as a whole is there ever certainty.

By an informal proof, Lakatos means the decomposition of the conjecture into a series of subconjectures making up a thought experiment that explains why the conjecture should be true.[1] Again, intended as characterizing the ordinary practice of proving a theorem, the subconjectures are usually formulated as a series of lemmas that together logically imply the conjecture's truth. Schematically, one might represent a conjecture C over some domain D of objects utilizing lemmas P_i as:

(*) $P_1(x)$ & $P_2(x)$ & . . . & $P_n(x) \rightarrow C(x)$ for all x in D.

Using Euler's $V - E + F = 2$ as C, one proof thought experiment might be the following series of lemmas, or subconjectures, formulated for a cube, the supposedly representative example used by Cauchy in his 1813 proof.[2]

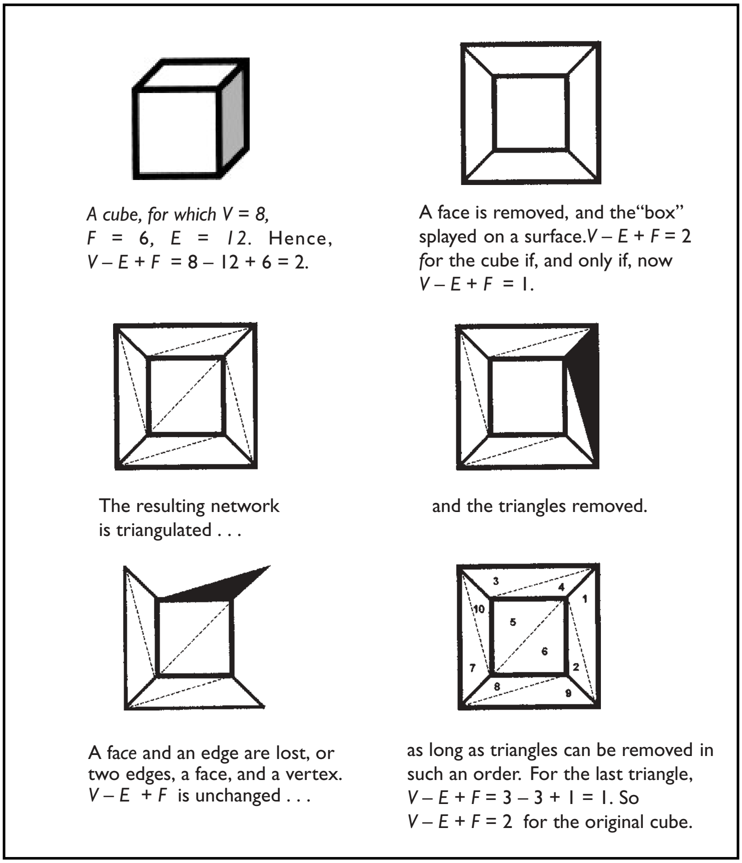


Figure 1. Proving Euler's theorem for a cube.

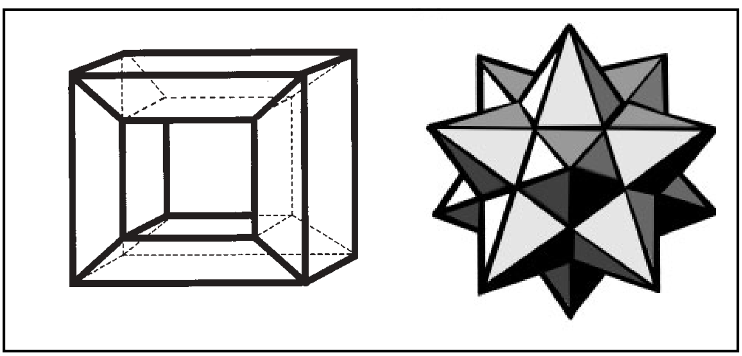


Figure 2. Picture frame and urchin polyhedra. The "picture frame" and "urchin" are counterexamples to some versions of Euler's theorem. The urchin's polygonal faces are shaded with different grays

P_1 A face can be removed from a polyhedron, leaving the number of vertices and edges unchanged so that $V - E + F = 2$ for the original polyhedron if, and only if, $V - E + F = 1$ for the new polyhedron.

P_2 After a face is removed, the polyhedron can be mapped from three dimensions into two dimensions, preserving the number of faces, edges, and vertices in the form of a planar network.

P_3 The faces of the planar network can be dissected using straight lines into triangles (for instance, a diagonal for a square leaves two triangles).

P_4 From the resulting planar network, edges and vertices can be removed so that either one loses one edge and one face, or two edges, a vertex, and a face. And thus, $V - E + F = 1$ for the original network if, and only if, $V - E + F = 1$ for the resulting network.

P_5 By removing edges and vertices from the planar network, one is left finally with an ordinary triangle.

P_6 Because for a triangle, $V = 3$, $E = 3$, and $F = 1$, then $V - E + F = 1$ also holds for a triangle, which implies that $V - E + F = 2$ for the original polyhedron from which one face was removed.

Much can happen with such a thought experiment, as summarized through the various heuristic patterns or shapes of mathematical reasoning provided in *Proofs.* Here is a brief summary of some of the heuristics presented there.

One may be confronted with a counterexample to *C,* say a "picture frame" (see figure 2), for which $V - E + F = 0$ (P&R 19). Lakatos calls the picture frame a "global counterexample" because it refutes the conjecture, not just a supporting lemma, and different responses to the global counterexample are illustrative of different heuristics for modifying the proof. The heuristic pattern of *monster-barring* to save the theorem here would mean denying that the picture frame is a polyhedron, though probably not denying it is a geometrical or mathematical object at all. Alternatively, *monster-adjustment* means, for instance, reinterpreting the configuration of edges and faces of, say, the "urchin" (see figure 2), for which $V - E + F = -6$. The shading shows twelve star-pentagon faces, with each pentagon having five vertices and five edges. Each vertex is common to five pentagons and each edge is common to two pentagons. Hence $V - E + F = 12 - 30 + 12 = -6$. If each star-pentagon is seen instead as being composed of five triangles, then there are sixty triangular faces, ninety edges, and thirty-two vertices. Recounting gives $V - E + F = 32 - 90 + 60 = 2$ for the newly "adjusted" urchin after all.[3] As another heuristic, returning to the picture frame, one might conjecture that it identifies *a,* and if one is lucky, *the* limited class of polyhedra D' for which the conjecture holds, say convex polyhedra, which excludes the picture frame. Here, the conjecture is adjusted by *exception-barring,* but the proof is left untouched: the modified conjecture is that for all *convex* polyhedra, $V - E + F = 2$, with the proof as before. One may also, in addition to restricting the domain D of the proof to objects D' excluding the counterexample, also inspect the proof to see just which lemmas are false for the counterexample, and revise or even replace them, so that in their improved versions they capture exactly what goes wrong with the proof for the counterexamples. For instance, lemma P_2 fails for the picture frame. Only those polyhedra that can be "inflated" into a sphere can be stretched onto the plane after removing a face, and the picture frame inflates to a torus or "doughnut." The picture frame is now seen as not only a global counterexample to the conjecture but a *local* counterexample to one of the lemmas as well. If we call polyhedra for which lemma P_2 is true "simple" polyhedra, then the revised conjecture is: for all *simple* polyhedra, $V - E + F = 2$. The difference from exception-barring is that the domain is limited to just the domain of the guilty lemma.

The virtue of such *lemma-incorporation,* for Lakatos, is that it improves the proof and conjecture by explaining why the original conjecture is false for the counterexample. To a first approximation, lemma-incorporation is presented in *Proofs* as the motor force of the method of proofs and refutations, superseding, through the voice of Lambda in Lakatos's reconstructed dialogue, monster-barring as often presented by Alpha, and the exception-barring of Beta and Sigma. Lemma-incorporation, unlike exception-barring, increases the rigor of the proof by building counterexamples into the lemmas and conjectures. It promotes identifying any hidden lemmas in the proof as well. Consider a cylinder, for which $V - E + F = 1$. It is a global counterexample, if it is considered to be a polyhedron, but why then does the proof fail? After removing the cylinder's "jacket," it can be stretched onto the plane, but this results in two disconnected disks; indeed, the proof depends on the planar network being *connected,* which therefore is a hidden lemma, and the proof again can be changed.[4] Before discovering the hidden lemmas, the cylinder was a global and nonlocal counterexample, while exposing the hidden lemmas turns it into a global and local counterexample, and proof revision can continue as before. Lambda, then, defines a theorem as "rigorous" when there are no such global and nonlocal counterexamples: "I call this criterion the *Principle of Retransmission of Falsity* [after Popper] because it demands that global counterexamples be also local: falsehood should be retransmitted from the naive conjecture to the lemmas, from the consequent of the theorem to its antecedent" (*P&R* 47). Thus, Lakatos provides an informal characterization of logical necessity—not according to the positive transmission of truth forward but the provisional absence of falsification backward.

Why should this process not seem completely obvious and trivial? How does lemma-incorporation amount to much at all? One answer, as amply documented in *Proofs,* is that what is obvious with the benefit of modern logical theory was not so during much of the nineteenth century. But still, why not just formulate heuristic revisions to the proof, perhaps with changes to C and domain D, as a change to its underlying logical form? That is,

$$(^{**})\ P'_1(x)\ \&\ P'_2(x)\ \&\ \ldots\ \&\ P'_n(x)\ \&\ P'_{n+m}(x) \rightarrow C'(x) \text{ for all } x \text{ in } D'.$$

The difficulty with (*) and (**) is that they suggest that the changes among C, D, and the P's can be "routed" through some prior and fixed formal language, including various geometrical transformations for the P's and some implicit universe of objects allowing various D's. But in revising the conjecture and proof from (*) to (**), the same rules might not work to translate (*) *into* the formalism for getting (**), as the revised proof and theorem, "out" of the formalism. Some of the P's in (*) might be false or contain hidden lemmas and themselves have to be replaced: what indicates, in advance, *which* formal replacements would then be selected for (**), and how would these be reexpressed in the language of informal mathematics? Also, if the P's are formally inconsistent, then any C follows as a consequence, and apparently the informal proof means nothing at all. Or if C turns out to have been false, then how could C be formally improved so that it is distinguished from any other formal inconsistency? For Lakatos, conceptual changes to the P's, C, and D, and the transition from (*) to (**), are only partially represented by their underlying formal logical structure. At the same time, Lakatos claims that the method of proof and refutations, largely lemma-incorporation plus the principle of the retransmission of falsity, drives an informal proof to rigorous, logical justification.

In Lakatos's lemma-incorporation, the eventual—conjectural!—logical truth of a theorem is won through the painstaking process of revising false proofs, and conversely these increases in rigor may also contribute to increases in content. But one may also never get out of falsehood. Improving a proof means reducing falsity, but unless one knows how to select a final formalization of set theory, or avoid the skeptical consequences of Gödel's incompleteness theorems, reducing falsity does not imply certainty. Reasoning about truth and falsity, in *Proofs,* is this many-layered process, and the error, as Hegel observes, is that truth and falsity are "considered immobile separate essences, as if one stood here and the other there, without community, fixed and isolated. Against this view one must insist that truth is not a minted coin which can be given and pocketed ready-made" (*PF* 58). The method of proof and refutations even suggests, as a heuristic, that you set out to prove *and* refute a conjecture, because it may be false. So you can make progress by "proving" a false conjecture. How might these mobile conceptions of truth and falsity and their ordinary logical interpretation be reconciled? The answer, in Hegel's words, is that "reason is purposive activity" (*PF* 34) as is proving theorems, but only one purposive goal is to justify a conjecture as true. Other goals are to discover new proofs; better organize a long sequence of progressive improvement; replace naive conceptions with a body of mature theory and axiomatics; criticize incorrect proofs and perhaps underlying mathematical standards; extend the domain of objects to which a proof applies; or create or discover novel mathematical objects. Truth and falsity are deployed ambiguously in Hegel and Lakatos because knowledge is seen as both product and process: the failures and successes of knowledge as goal-seeking process become truth or falsity only in the static theorem product.

The method of proof and refutations does not reject modern logical form but states the heuristic logic used in moving back and forth between informal and formal justification without assuming the latter's prior existence or validity. One point of the historicism in *Proofs* is to explain progress in standards of rigor during the nineteenth century, but without assuming the fully formed modern notions of logical truth and formalization, which emerged from a process through which the boundaries of mathematics and patterns of mathematical reasoning themselves changed. Ultimately, it is that kind of major conceptual change that Lakatos wants to capture through lemma-incorporation, or the renamed method of proof and refutations to be changed even later to proof*s*. While the historical changes in proofs of Euler's theorem included changing conceptions of polyhedra, and three- and two-dimensional geometry, they were not mathematically revolutionary. But in the same way that Euler's theorem introduced new geometrical objects into mathematics, other parts of mathematics, almost primarily those involved with real analysis or functions of real numbers—calculus, in short—were grappling with such novelty on the boundary of mathematics throughout the century. A fixed formal representation, and a fixed universe of mathematical objects, is going to set some limits for what mathematics includes, or how one reasons within it. But there is no noncircular way to set those limits so that the important transformations in nineteenth-century mathematics—involving changing notions of functions, sets, series, continuity, dimension, and area—are "allowed" as changes from a formal representation in which they did not yet occur. Popper once suggested an analogous example in that one could not have predicted the invention of the wheel in primitive times, for a necessary part of that would be to say what a wheel is, and to say that would be to invent the wheel itself. "Ordinary" lemma-incorporation—such as moving about the boundary of "polyhedra" among various two-dimensional surfaces embedded in three-dimensional space, or reinterpreting the picture frame or urchin—when represented as the change from (*) to (**) is correct today, but historically misleading and incomplete from a fallibilist philosophical perspective. The potential for modest conceptual change is at the heart of lemma-incorporation, and lemma-incorporating revisions to most mathematical conjectures do not require sweeping changes to mathematics as a whole. But they *might,* and according to Lakatos, they did lead to novel forms of mathematical knowledge in the nineteenth century, and even in part to modern formalisms.

So far, there is a close fit between the overall historiography and the philosophy of learning through error in both *Proofs* and the *Phenomenology.* Lakatos's empirical conception of informal, purposive proof, and informal truth and falsity, are consistent with Hegel's conception of rationality as process. The next step is to see that these affinities go a level deeper, to the *Phenomenology*'s historiographical method and Hegel's so-called method of phenomenological criticism. Toward that goal, four issues related to Lakatos's heuristics and his historiography shall serve as guides.

First, how is lemma-incorporation, or the method of proofs and refutations, related to Lakatos's historiography? What are Lakatos's significant historical claims, and how does his historiography help elucidate their philosophical implications? Marx and Hegel make specific claims about the history of capital and social institutions; what does Lakatos say about the history of mathematics?

Second, what is conceptually Hegelian about the method of proofs and refutations, over and above the bildungsroman form for Lakatos's historical reconstruction? As we shall see, lemma-incorporation is a carefully articulated and logically coherent version of that aspect of Hegel's phenomenological criticism by which one criterion or standard of knowledge is improved to another by a change brought about by an object of knowledge in conflict with the former standard, but not the latter. For Lakatos, the "objects" of knowledge are non-Platonic mathematical entities, and the "criterion" is a theorem and its proof. Hegel's phenomenological criticism, stated briefly, is a historiographical technique for characterizing patterns of knowledge, including the history being constructed, such that: what is described are relations among the criteria or standards of knowledge and their intended objects; the criteria characterizing a shape of knowledge—such as Kantian Verstand, or Sense-Certainty, or the Unhappy Consciousness, or others like Lakatos's caricatured shapes of mathematical knowledge—are immanent in its expression, and standards of philosophical criticism devolve from the historical expression of knowledge itself, allowing us, as Hegel alleged, simply to "look on" at these historical "appearances"; what we "observe" includes contradictions between criteria of knowledge and their intended objects, with progress characterized as the resolution of such contradictions in a new, superseding standard. Lakatos has distilled a core from Hegel's methodology that, rather than describing changing patterns of consciousness, depicts historical changes in terms of changing languages for mathematical theorems and their proofs.

Third, a key element in Lakatos's account of the importance of lemma-incorporation for contemporary mathematics, and one that ties his philosophy of mathematics to his historiography, is the role of skepticism in informal mathematical method. Hegel describes his phenomenological method via the methodological tropes and concepts of ancient Pyrrhonian skepticism. Lakatos also turns methods of classical skepticism into the principal conceptual mechanism underlying his learning through error. But unlike Hegel, Lakatos sees a potential for epistemological chaos in reason owed to this very presence, and he characterizes the crisis in mathematical foundations at the turn of the nineteenth century as mathematical knowledge approaching, and then withdrawing from, such a skeptical abyss. The historical problem of the foundations of mathematics, therefore, is characterized as having at its heart a fundamentally skeptical conception of mathematical progress. The skeptical and antifoundational lessons of Gödel's incompleteness theorems, as we shall see, on this account have a deep historical rationale. Methodological skepticism unifies Lakatos's historiographical portrayal with both the method of proofs and refutations and his history of the mathematical present.

Finally, since the historiographical style of *Proofs* creates a philosophical history of the mathematical present, the method of proofs and refutations should be implicated in the emergence of the logical and metamathematical standards of rigor familiar today. Lakatos outlines how formal criteria for mathematical truth grew out of lemma-incorporation as a new "shape" of mathematical reasoning, and the role here of lemma-incorporation is in the translation of informal proofs into formalized mathematics. This procedure will be illustrated in chapter 4 through the role of lemma-incorporation in Gödel's second incompleteness theorem and the translation of informal consistency into a formalized consistency statement. Gödel's second theorem thus emerges in a new historical light: the theorem states one of the most striking skeptical consequences of mathematical logic, namely, the unprovability of consistency, but it is also a historically representative theorem in that its mature statement required the process of proofs and refutations acting explicitly between informal and formal mathematics. Gödel's second theorem, in this way, is symptomatic of changes in mathematics over a century ago, and its curious consequences may be indicative of an older and persistent logic underlying its surface form.

Dirichlet, Cauchy, and Seidel

Lakatos's repeated historical claim is that lemma-incorporation was a new feature of mathematical reasoning that emerged full force near the middle of the nineteenth century. Such proof analysis, he says, was

> virtually unknown in the informal mathematics of the early nineteenth century [and] . . . in the mathematics of the eighteenth century. . . . [It] is no coincidence that the discovery of the method of proof and refutations occurred in the 1840s, when the breakdown of Newtonian optics (through the work of Fresnel in the 1810s and 1820s), and the discovery of non-Euclidean geometries (by Lobatschewsky in 1829 and Bolyai in 1832) shattered infallibilist conceit. . . . Why did the leading mathematicians from 1821 to 1847 fail to find the simple flaw in Cauchy's [continuity theorem] proof and improve both the proof-analysis and the theorem? The first reply is that they did not know about the method of proof and refutations. They did not know that after the discovery of a counterexample they had to analyze their proof carefully and try to find the guilty lemma. They dealt with global counterexamples with the help of the heuristically sterile exception-barring method. (*P&R* 48n1, 60n1, 139, 136)

Lakatos supports this historical hypothesis by isolating and analyzing intermediate stages of rigor emerging in the mid–nineteenth century that advance beyond those found earlier, notably Cauchy's exception-barring, over which lemma-incorporation is reconstructed as the superior shape of mathematical knowledge. Exception-barring, claims Lakatos, also falls short of what is found in later-nineteenth-century mathematics after Karl Weierstrass and Bernhard Riemann. What is the "actual history" underlying Lakatos's claims?

In 1848, Peter Lejeune-Dirichlet's student Philip Seidel published a note on series of functions representing discontinuous sum functions. The paper began by citing the infamous false "continuity theorem" from Cauchy's pathbreaking *Cours d'analyse algebrique* of 1821, which incorrectly proved, along with other erroneous theorems, that the sum of a convergent series of continuous functions is itself a continuous function. That is, if each $f_n(x)$ is continuous on an interval (a,b) containing x and $F(x) = \Sigma f_n(x)$ is a convergent series on (a,b), then Cauchy claimed that F is continuous at x also. Counterexamples to Cauchy's continuity theorem, and other false theorems of the nineteenth century, were not apparent, and as a result, this false theorem and others persisted through the century, consistent with Lakatos's view in *Proofs.* Figure 4, in the appendix to this chapter, is an example of convergent series for which Cauchy's theorem fails, and with the converging sum functions $f_n(x)$ also meeting the conservative definition of continuity assumed by the *Cours.* After citing Cauchy's theorem, Seidel continued:

> Nevertheless, the theorem stands in contradiction to what Dirichlet showed, that, for example, Fourier series also always converge if one forces them to develop discontinuous functions;—indeed, the discontinuity will frequently be embedded in the form of these series whose individual members [that is, sine and cosine functions] are still continuous functions.[5]

Seidel then stated and proved a theorem that is recognized today as a special case of Weierstrass's later correct theorem that the *uniformly* convergent sum of continuous functions is again itself continuous.[6]

Cauchy's false continuity theorem, published twenty-seven years before Seidel's note, had been one focus of a debate reaching back into the eighteenth century over the limitations of representing functions by trigonometric series, and especially by the controversial Fourier series, through which a large class of functions may be unintuitively expressed as the infinite sum of simple trigonometric functions. In 1829, Dirichlet published his famous proof, where Fourier, Cauchy, and Poisson had failed, finally establishing the convergence of Fourier series for functions meeting what are now called the Dirichlet conditions.[7] The generality of this theorem was a remarkable achievement, showing that nonperiodic functions, and especially discontinuous functions with "jumps," could be expressed as the infinite sum of periodic continuous functions. These sophisticated infinite series were the ones raised by Seidel as counterexamples to Cauchy's publication over two decades earlier. Though Dirichlet's theorem clearly showed that Cauchy's theorem must be false, and while Cauchy's theorem was not generally considered completely sound, it was not until Seidel's note that the contradiction between Cauchy's and Dirichlet's work was made explicit and finally resolved.

That Cauchy's false theorem at least "admits of exceptions" was known before 1829, when Dirichlet published his proof, and it was also understood that the counterexamples could be produced via infinite series like those considered by Dirichlet—that is, Fourier series converging to discontinuous sum functions. In 1826, Niels Abel, who appears with Cauchy in *Proofs* as another exception-barrer, attached a footnote to a paper on binomial series in which he remarked "that [Cauchy's] theorem admits of exceptions," and "there are, as one knows, many series of this type."[8] Abel apparently did not draw the conclusion that Cauchy's theorem was explicitly falsified by his counterexample, in today's sense. Rather, he discovered problems of *conditional* convergence, in which some series may be rearranged so as to converge or not, and then to different values. Abel's response to conditionally convergent series was, true to the exception-barring spirit, to exclude them entirely, an attitude that persisted until the 1890s. And in examining Cauchy's false continuity theorem, Abel did not see how it was driven by the pointwise versus uniform convergence mode. He and others stopped short of exploring the properties of a whole class of series of great mathematical value. In this way, exception-barring retards mathematical progress. Cauchy, for his part, appeared to be unaffected by it all; he seemed aware of problems with his theorem on publication, and in 1833, four years after Dirichlet's publication, published a set of lectures in which he repeated his theorem and its proof verbatim from the *Cours d'analyse.*[9]

Lakatos's characterization of Abel and Cauchy as methodologically inferior exception-barrers, then, is historically correct. So, too, is the novelty of Seidel's methodological criticism vis-à-vis the persistent problems of early-nineteenth-century analysis and Cauchy's partial progress. In *Proofs,* Seidel represents a crucial methodological juncture in nineteenth-century mathematics, standing between the level of rigor represented by Cauchy, Abel, and others' admitting of exceptions, and the Seidel-style reformulation of Cauchy's continuity theorem to exclude and explain the counterexample. Seidel, in spite of his lesser fame, is identified in *Proofs* as articulating the method of lemma-incorporation, and he does so quite clearly in his criticism of Cauchy.[10] Thus emerges one rationale for Lakatos's intricate historiography: while the theorem debated throughout *Proofs* is nominally Euler's, the original proof is Cauchy's, and the methodological criticism epitomizing the problems of Cauchy's proof belongs to Seidel, the latter being embedded in a controversy of real analysis only indirectly related to the polyhedra relevant to Euler's theorem. One could conceive a version of *Proofs* oriented around changes in the function concept and its nineteenth-century odyssey, but such a work would lack the perspicuity of a simpler "generic" theorem like Euler's. At bottom, though, the methodological innovations are the same.

Just as in *Proofs,* where the ongoing story involves what should count as a polyhedron, the parallel argument here in real analysis, which was to change mathematics entirely, was over the status of mathematical *functions* generally. Euler had formulated a general definition as early as 1748 that accommodated both "analytic" functions, defined using a single explicit formula, and the modern notion of an arbitrary correspondence that is not necessarily specified by a single algebraic rule, like $y = x^2$.[11] But in practice, for much of the nineteenth century, relations not described by an analytic expression were *not* considered functions, and even more conservative definitions were used for continuous versus discontinuous functions, as compared to present practice. Cauchy, for one, may have not intended his conception of continuity to include functions with corners, like figure 3(*b*), exactly the kind of function continuity is supposed to capture today. Fourier series, therefore, implicitly raised questions about just what continuity, functions, and convergence all meant, and much of it on the boundary limiting what was mathematically allowable at the time. The controversy was not resolved with Cauchy, or Dirichlet, but continued throughout the century. Cauchy's false continuity theorem was one such locus of contention, and the theorem existed in a kind of netherworld of truth contaminated by exceptions, quite unlike our contemporary standards. These truth standards shared in both the haphazard standards of Cauchy's eighteenth-century predecessors, whom he started to criticize so well, and the standards of rigor yet to be introduced by Weierstrass's school, which would still need to correct Cauchy, his many achievements notwithstanding. In spite of Cauchy's critical advances making possible Weierstrass's and our own conception of real analysis, Cauchy retained notions of mathematical truth, falsity, and validity that, with Seidel, finally began to be regarded as generally inadequate. Cauchy is wrongly credited in some histories of mathematics for being more rigorous than he was and creating a type of mathematical truth that only half existed then.[12] This is Lakatos's historical context for the emergence of lemma-incorporation.

Lakatos's perspective, in contrast to many accounts of nineteenth-century mathematics, is not one about the specific content of important developments, such as the elimination of infinitesimals and the arithmeticization of analysis. The difference between Cauchy and Weierstrass on this account is not, for example, that Cauchy's analysis maintained key ambiguities of infinitesimals—say, whether "*dy*/*dx*" is to be taken as a unitary symbol denoting the derivative, instead of *dy* and *dx* having independent status as vanishingly small quantities—while Weierstrass clearly saw the need to eliminate infinitesimals. Nor does Lakatos's thesis depend on submerged issues that did not begin to be resolved until the work of Georg Cantor, Richard Dedekind, Gottlob Frege, and Giuseppe Peano in the latter half of the nineteenth century, including changing theories of real numbers. In Lakatos's historiography, then, the change from the early-nineteenth-century view that logic was irrelevant to mathematics is made into a protracted theoretical development only started by Cauchy and including Seidel's criticism as an intermediate step.[13] Seidel's statement of lemma-incorporation and the method of proofs and refutations represents a changing standard of rigor in which "exceptions" came to be taken as counterexamples, thereby making necessary a reformulation of proofs and concepts not characteristic of earlier years.

Subject and Object, Theorem and Proof

Lakatos's basic historical thesis about the novelty of lemma-incorporation, then, is well grounded. Lemma-incorporation is also a species of falsification. But to return to our Hegelian interests, what is it about the method of proofs and refutations that may give it more than an affinity to post-Kantian philosophical notions of error and criticism?

The weakness, for Lakatos, of the earlier exception-barring method lies in the separation of the theorem from its proof. For example, convexity of a polyhedron is nowhere used in Cauchy's proof of $V - E + F = 2$. It is also not hard to see that Cauchy's proof does work for spheroid polyhedra, with indented, concave faces, such as a globe with a political map drawn on it, so that convexity has nothing to do with *this* proof. Nevertheless, the restriction of Euler's theorem to convex polyhedra was a common means of ensuring the validity of the theorem in the face of nonconvex counterexamples for which Cauchy's proof failed, but not because of nonconvexity. Convexity is a good guess at the domain of the proof and excludes many refuting counterexamples, although as Lakatos maintains, "It forgets about the proof; in guessing the domain of validity of the conjecture, you do not seem to need the proof at all" (*P&R* 29). The exception-barring method does not explain *what a* *proof proves,* meaning the reflective question of why the restrictions imposed by the conditions of the theorem are necessary to make the proof work, and why the proof, yet perhaps not a different proof of an equivalent theorem, breaks down otherwise. Remarkably, this commonplace practice of contemporary mathematics was not characteristic of the early nineteenth century. Just identifying convexity does not incorporate the counterexample into the proof as a lemma, condition, or revised definition; instead, as with the most "primitive exception-barrers, the proof and the exceptions exist in two completely separate compartments" (*P&R* 36). As Lakatos's Teacher describes "the most important aspect" of lemma-incorporation:

> The casual statement that in your restricted domain all the lemmas will be true whatever they are is enough for your purpose. But this is not enough for mine. I build the very same lemma which was refuted by the counterexample into the conjecture, so that I have to spot it and formulate it as precisely as possible, on the basis of a careful analysis of the proof. The refuted lemmas thus will be incorporated in my improved conjecture. Your method does not force you to give a painstaking elaboration of the proof, since the proof does not appear in your improved conjecture, as it does in mine. . . . [False] proofs, even though they may not prove, certainly do help to improve our conjecture. *The exception-barrers improved it too, but improving was independent of proving. Our method improves by proving. The intrinsic unity between the "logic of discovery" and the "logic of justification" is the most important aspect of the method of lemma-incorporation.* (*P&R* 36–37)

"Intrinsic unity" means that the counterexample has been used to substantively improve the false conjecture into a new one so that the increased rigor, logical validity, or justification is structurally embedded in the revised lemmas or definitions. It is also just the type of change that Hegel asserted constitutes progress between his reconstituted patterns of knowledge in the *Phenomenology.* Intrinsic unity in the phenomenological method is the incorporation of criticism into a shape of knowledge, so that, as Hegel argued,

> in every case the result of an untrue mode of knowledge must not be allowed to run away into an empty nothing, but must necessarily be grasped as the nothing *of that from which it results*—a result which contains what was true in the preceding knowledge . . . only when [a new form of knowledge] is taken as the result of that from which it emerges, that it is, in fact, the true result . . . which has a content. (*PhS* 51)

Or stated another way by Hegel:

> The exposition of the untrue ["falsified"] consciousness in its untruth is not a merely *negative* procedure. The natural consciousness itself normally takes this one-sided view of it; and a knowledge which makes this one-sidedness its very essence is itself one of the patterns of incomplete consciousness [like exception-barring] which occurs on the road [to Hegel's Science] itself. . . . This is just the skepticism which only ever sees pure nothingness in its result and abstracts from the fact that this nothingness is specifically the nothingness *from which it results.* For it is only when it is taken as the result of that from which it emerges, that it is, in fact, the true result; in that case it is itself a *determinate* nothingness, one which has a *content.* . . . But when . . . the result is conceived as it is in its truth, namely, as a *determinate* negation, a new form has thereby immediately arisen, and in the negation the transition is made through which the progress through the complete series of forms comes about of itself. (*PhS* 56)

I prefer Lakatos's formulation of the separation of counterexamples and their incorporation into falsified lemmas. But the unity achieved through lemma-incorporation is an instance of just this type of progress identified for Hegel's shapes of consciousness, in which errors are reformulated and intrinsically improved in the succeeding shape of knowledge. One difference with Hegel is striking: while a new shape of knowledge is reconstructed as progress for Hegel, but not necessarily as it occurred historically, for Lakatos, it is the mathematical practice of such "retransmission of falsity" that helped generate a cornucopia of historical theorems, not just, as in Hegel, a self-contained series of types of knowledge. The later-nineteenth-century refinements of Weierstrass, being the mature practice of lemma-incorporation, in Lakatos's view "united logic and mathematics for the first time" (*P&R* 55), and it came to be that as a general feature of mathematical method, counterexamples were finally neither listed as exceptions nor segregated from the proof by a safe guess that may have had little to do with the proof itself. Instead, the Weierstrassians "brought about the integration of proof (thought experiment) and refutations and started to develop proof-analysis, gradually introducing deductive patterns in the proof-thought-experiment" (*P&R* 55). For Lakatos, it was Weierstrass's school that ultimately practiced lemma-incorporation with systematic fervor, thereby redefining standards of mathematical rigor, and thus helping to make possible the abstract ideas of proof familiar today.

This proof analysis is not the symbolic formalization introduced toward the end of the nineteenth century by Peano but rather the informal interpretation of a proof that makes such formalization possible at all. At the same time, this methodological change in *Proofs* is the same as that part of the phenomenological method through which one pattern of reconstructed knowledge progressively succeeds another in Hegel's historiography. In both lemma-incorporation and the phenomenological method, the change to a new shape of knowledge implies, as Hegel noted, that "the criterion for testing is altered when that for which it was to have been the criterion fails to pass the test; and the testing is not only a testing of what we know, but also a testing of the criterion of what knowing is" (*PhS* 54); for lemma-incorporation, it is the theorem at stake as the criterion of mathematical knowledge that has changed, with a retransmission of falsity back to the theorem's lemmas. It is in this intrinsic unity of a counterexample with a superseding theorem, not achieved by monster- or exception-barring, that lemma-incorporation, as a reflective understanding of the proof, becomes the important historical practice first articulated by Seidel in championing the implications of his mentor's work on Fourier series. No doubt, Lakatos has a clearer conception than Hegel of how this process is related to classical logical justification. But for nonformalized mathematical discourse, Lakatos's crucial intrinsic unity within a change of knowledge is pure Hegel. To see further how Lakatos has implicitly transformed his phenomenological ancestor, it is helpful to inspect lemma-incorporation more closely.

Toward that goal, we turn to the character Pi, Lakatos's spokesperson for skepticism in the foundations of mathematics, and more a philosopher of language than mathematics. Pi is first introduced in a discussion of how "different proofs yield different theorems" (*P&R* 66), and most of what Pi says is an elaboration of that idea. Pi's insight into multiple proofs again is not timeless, though it may appear so when presented as heuristic advice. As Salomon Bochner says of the publication of Carl Friedrich Gauss's *Disquisitiones Arithmeticae* in 1801,

> The nineteenth century in mathematics arrived in earnest somewhat later, when Gauss saw himself "motivated" to produce four different proofs for the assertion that a polynomial with complex coefficients has a complex root. It is reported that the aging [Joseph Louis] Lagrange was puzzled by this attitude; from his eighteenth-century approach it was indicated to have just one proof, a satisfactory one, and to continue from there with something else.[14]

This methodological idea, though commonplace today, was relatively novel in the early nineteenth century, and Lakatos's investigation is about what occurs conceptually through what is often a formally uninformative process.

Pi begins by noting how different proofs of Euler's theorem change the theorem because they allow the proof to be applied to different classes of mathematical objects. For example, the proof can be transformed into a counting rule for calculating the value of $V - E + F$, whatever it turns out to be, for all kinds of polyhedra, for example, $V - E + F = 0$ for a picture frame. Pi explains that when a proof is modified to account for falsifying counterexamples, the theorem is no longer just about polyhedra, but different mathematical entities associated with the eponymous proofs to which they apply. As Pi claims, "The theorems are about Cauchy-objects, Gergonnian objects, Legendrian objects respectively, but not any more about polyhedra" (*P&R* 66), meaning those various geometric objects to which the respective proofs may be applied, regardless of the original naively defined domain of polyhedra: polyhedra with cavities or different connected structures, higher-dimensional surfaces, and with Poincaré's modern proof discussed at the end of *Proofs,* even classes of abstract vector spaces. What were originally fringe objects, nonpolyhedra, or counterexamples are now taken as what *should* be known via the proof, instead of upholding the theorem as what *should* be true. The different proofs allow the original conjecture to be about, or extended to, very different objects than intended originally, and explain their properties, but not exactly as expected—for example, by accepting that $V - E + F$ may differ from 2 for interesting reasons. As Hegel put it, still describing his phenomenological method, "In the alteration of the knowledge, the object itself alters for it too. . . . [A]s the knowledge changes, so too does the object, for it essentially belongs to this knowledge" (*PhS* 54–55). That expresses a general theory-observation dependency between the subject and objects of knowledge for which Hegel is well known, and which Lakatos now identifies through the changing and stretched concepts of mathematical proofs.

Pi fills out the enigmatic comment about "different theorems" some seventeen pages later, during the exposition of mathematical concept formation. Recalling the early pages of *Proofs* and the methods of monster-barring, monster-adjustment, exception-barring, and even lemma-incorporation, Pi turns the entire account developed so far on its head by challenging the inferior status of the unsophisticated monster-barrers vis-à-vis the proof and refutationists. The monster-barrers refuse to recognize stark counterexamples to Euler's theorem such as a picture frame (for which $V - E + F = 16 - 32 + 16 = 0$)

as a polyhedron at all, thus preserving the truth of the theorem by circumscribing its domain post hoc, and classifying as pathological "monsters" the nonpolyhedral counterexamples Cauchy or others might label "exceptions." Until Pi's emergence in the dialogue of *Proofs,* the monster-barring antifalsificationists are portrayed as restraining the progress of mathematics facilitated by lemma-incorporating criticism. Pi raises the possibility that the monster-barring refusal to accept a counterexample such as the picture frame as a genuine refutation, and not cleverly contrived pathology, may not necessarily indicate the cowardly contraction of the theorem's domain to a safe set of objects. It may just as well indicate surreptitious or unconscious *concept stretching* on the part of refutationists, who willy-nilly expand the intended domain, lacking as it may be in correct definition, and thereby allow the falsification of the theorem through a counterexample where none existed before.

Before continuing with Pi's account, it is again important to mention the place of concept stretching in the mainline of nineteenth-century mathematics. For instance, in his celebrated paper establishing the convergence of Fourier series, Dirichlet also is sometimes credited with rediscovering Euler's general definition of a function as an arbitrary correspondence because of his example there of the characteristic function of rational numbers: $\boldsymbol{Q}(x) = 1$ if x can be expressed in the form p/q with p and q integers, and $\boldsymbol{Q}(x) = 0$ otherwise. So, for example, $\boldsymbol{Q}(.5) = 1$, $\boldsymbol{Q}(-100.1) = 1$, $\boldsymbol{Q}(.3333\ldots) = 1$, and $\boldsymbol{Q}(\pi) = 0 = \boldsymbol{Q}(\sqrt{2})$, since π and $\sqrt{2}$ are both irrational numbers: they cannot be expressed in the form p/q. At the time, it was believed that $\boldsymbol{Q}$ could not be expressed in a closed analytic form, which was surprisingly recognized to be false by Hermann Hankel in his 1870 critique of function definitions. Before then, $\boldsymbol{Q}$ was thus a good example of a nonanalytic definition, one not following an equational formula, and not to be considered a function at all, even though it is taken as one today. It also may be that Dirichlet did not want to include $\boldsymbol{Q}$ as a legitimate function; instead, he may have seen it as a definition going outside the function boundary.

What motivated that discipline-wide exclusion? Until about 1870, the goal for many mathematicians was to organize real analysis around "well-behaved" classes of functions with shared properties: continuous and differentiable functions, uniformly convergent functions, integrable functions, Fourier-expandable functions, and so on. The perspective of taking as valuable in their own right counterexamples or other monstrous objects not captured by the implicit intent of real analysis theorems was not generally held. Perhaps also in part due to the close association between mathematics and physics, rules like that defining the pathological ***Q***, which lacks useful properties and is not obviously found "in nature," did not define functions. Only later concept-stretching expansions, especially by Riemann, would make such fringe objects a part of received mathematics, meaning parts in which theorem proving and theoretical elaboration occurs for these entities. So while Dirichlet restated the idea of functions as general or arbitrary correspondence, in practice, none of his work made use of it; his definition served to delimit a boundary, not extend one, at least for some decades. As another example, Cauchy appears not to have intended his definition of continuity to be as extensive as the modern interpretation, which includes functions with corners, but believed that he was just reformulating the old Eulerian-analytic idea of a function in arithmetic terms.[15] The useful functions tacitly excluded as "noncontinuous" would appear later when theorem proving could catch up to them. In any case, in the informal mathematics of the nineteenth century, it was not possible to consistently pin down the referents of all mathematical terms with the precision of formal logic. The twin functions of concept stretching and monster-barring contraction were, as Lakatos indicates, important methodological tactics through which arguments about intended domains went hand in hand with the criticism and improvement of mathematical proofs.

In *Proofs,* as Pi says, this means that it could be that the original "conjecture [that is, Euler's theorem] was true in its intended interpretation, [and] it was only false in an unintended interpretation smuggled in by the refutationists. Their 'refutation' revealed no error in the original conjecture, no mistake in the original proof: it revealed the falsehood of a new conjecture which nobody had stated or thought of before" (P&R 84–85). From Lakatos or Lambda's perspective, the point is to explain that original "intention," and that articulation occurs through lemma-incorporation. Therefore, asserts Gamma, the monster-barrer's sequence of definitions all "denoted the same good old concept of polyhedron he inherited from his forefathers. *He defined the very same poor concept in increasingly rich theoretical frames of reference, or languages*" (P&R 86). That is, counterexamples sometimes only become refutations when accompanied by the informal expansion of the concepts defining the theorem's domain: this is Pi's surreptitious concept stretching. The counterexamples are formal, logical refutations in the expanded domain, perhaps also with an implied expansion to a new theoretical language. If the expanded domain is not recognized as legitimate, then with respect to the "intended," smaller domain, no refutation occurs, and this is Pi's new alternative to defend the monster-barrers: their refusals to deal with the monsters can be rationalized. But if the monster-barring intention "does not form concepts but only translates definitions" (ibid.), the characteristics of the new objects discovered by concept stretching, such as the pathological functions or weird polyhedra, are incorporated neither into the structure of the proof nor a conception of the theorem's domain. There is nothing "wrong" with an intended domain. But if the intended domain is defended with neither a good reason nor theoretical articulation of the domain, then the refusal to acknowledge concept stretching amounts to rejecting a heuristic counterexample if not a logical one, since the expansion of the domain of polyhedra or functions itself may involve an informal change in reference or other semantic relations. Euler's theorem *can* just be intended to be about a "fixed" class of polyhedra excluding ones with tunnels or cavities, and it is a concept-stretching expansion of this intended domain that creates some of the possibilities for falsification. The objects of mathematical knowledge, in this way, can be unstable vis-à-vis the theorems speaking of them. Pi exposes the critical choice of whether and how such object changes are addressed in the theorems and proofs from which they arose, with Abel's divergent series, Dirichlet's nonfunction ***Q***, and Cauchy's limited class of continuous functions all playing this role during the nineteenth century. With Pi's alternative defense given for the monster-barrers, the question for *Proofs* now is to distinguish critical progress from regress.

For Hegel, critical regression occurs in the *Phenomenology* when a standard of knowledge and its objects are kept apart, just as for the exception-barrer, the proof is segregated from its counterexamples. With the recognition of the ambiguities introduced by Pi's account of concept stretching, even if a theorem is conceived as "correctly" applying to an initial intended domain such as simple polyhedra, an improved proof may apply to a broader class of objects through a changed frame of reference or new theoretical language for the mathematical domain. That is, a proof may be improved by *either,* as in the pre-Pi formulation of lemma-incorporation, as Lakatos explains, implicitly stretching, not just generalizing, the domain and incorporating the new characterization of the domain into the proof as a condition; *or* taking the theorem's domain as fixed, possibly ruling out counterexamples by additional lemmas or conditions, but *also* discovering that the theorem or one like it can be applied to other kinds of mathematical objects.

This second role of learning about newly conceptualized mathematical objects provides a fresh venue for progress, even if the original conjecture is left unrefuted and made safe from monsters. Instead of insisting on changes to a conjecture's proof as constituting improvement, changes can be embedded in more sophisticated definitions of what the theorem is about. In this case, again analogous to or even the same as Hegel, the class of objects corresponding to the proof becomes the standard of truth in critical progress: "If we designate knowledge as the Concept, but the essence of the True *as what exists, or the object,* then the examination consists in seeing whether the Concept corresponds to the object," as opposed to "whether the object corresponds to its Concept" (*PhS* 53). Pi's account of surreptitious concept stretching in a theorem's domain is Lakatos's description of what occurs when the "object" side of the theorem-domain relationship becomes the epistemic standard to which the theorem should conform. But finally, the "direction" of criticism is immaterial, if not for later materialist critics of Hegel, then at least from Hegel's perspective of philosophical history or a history of mathematical ideas: "It is evident, of course," said Hegel, "that the two procedures are the same" (ibid.). Whether through subject- or object-oriented criticism, fundamental improvement in a theorem, for Lakatos, takes place precisely as in phenomenological criticism: by lemma-incorporating counterexamples or through more sophisticated, stretched conceptions of a mathematical domain into the next shape of knowledge. This central aspect of Hegel's phenomenological method is translated by Lakatos not just into an account of nineteenth-century mathematics but into a feature of mathematical criticism itself, as a convolution of subject-theorem and object-domain. What is happening in that process?

The complete process of proofs and refutations, meaning the transition from a theorem to its improved version via lemma-incorporation, cannot necessarily be modeled, according to Lakatos, in a fixed formal language—not because it is historical per se but because concept stretching can change the underlying minimal languages necessary to represent the semantic changes taking place: there is no definitive means for identifying *the* true intended domain of a theorem or *the* boundaries of mathematics as a whole. "This theory of concept-formation," writes Lakatos about Pi's account, "weds concept formation to proof and refutations" (*P&R* 90n1), in that how counterexamples are managed with respect to a body of theorems directs a larger understanding of whole mathematical domains. For Lakatos, the preferred direction is when knowledge expands by discovering the boundary to something new. At such points in mathematical learning, "one does not prove what one has set out to prove," or "one does not solve the problem one has set out to solve," partly because in the generation of a new proof, "your term no longer denotes what it set out to denote" (*P&R* 90). Some of the most significant mathematical falsifications for Lakatos are those in which original intent is foiled and allowed to be redirected through changing referents instead of improved proofs; that is Pi's answer. Or as put more generally by Hegel regarding the transformation of knowledge: "One learns that what one supposed was not what one was supposed to suppose" (*PF* 97), and it is those experiential lessons that become historical learning and progress.[16]

The play on words refers simply to the several ways in which learning can occur. Pi's general notion that different proofs prove different theorems means that the errors of an original naive conjecture can be realized in unanticipated, and perhaps historically novel, domains and theoretical terms. What a proof proves may then include the creation of a changed conceptual, theoretical, and linguistic setting in which the theorem is cast. In the case, say, of Dirichlet's ***Q*** or other nineteenth-century monsters, the new language needed to conceptualize arbitrary correspondence was the theory of point sets and relations. The relation of the old proof to the new one, the old descriptions to the new ones, the old objects to the new ones, and the old theorem to the new one is that the naive, ungebildet concepts with which Euler's theorem began have now disappeared and been replaced with proof-generated concepts that, as Lakatos observes, "are neither 'specifications,' nor 'generalizations' of naive concepts. The impact of proof and refutations on naive concepts is much more revolutionary than that: they erase the crucial naive concepts completely and replace them by proof-generated concepts" (*P&R* 89–90).

This characterization by Lakatos of the changes brought about by lemma-incorporation falls just perfectly short of Hegel's famous notion of *aufheben,* which unites both antithetical meanings of the word: supersession as both cancellation or negation and preservation or affirmation. This dual sense for *Aufhebung* is intended by Hegel to describe the progressive learning by consciousness as shapes of knowledge are enumerated in the *Phenomenology.* The term is justified by Hegel's depiction of consciousness as depending on criteria of knowledge, "counterexamples" to the criteria, and their improved, "lemma-incorporated" successors. Progress is real for Hegel because changes in knowledge become embedded, materially or otherwise, in successive shapes of knowledge or the ways of life expressing them. But the analogous transformations among mathematical theorems must just miss such sublation, for the developments Lakatos analyzes fall within his rational reconstruction of successive proofs, and are not to be found in single concepts isolated from their history, or even the process of actual discovery: there is no Aufhebung as residue, so to speak, in the proof itself.[17] Historical progress is embedded in the proof only insofar as *we* see it as culminating its particular history: historical change, while perhaps constitutive of human or social consciousness as in the *Phenomenology,* is not intrinsic to mathematical proof per se. Human consciousness can be said to "be" historical in a strong ontological sense that does not translate well into mathematical proof. In spite, then, of the intrinsic unity of the logic of both discovery and justification, as Lakatos points out, "the zig-zag of discovery cannot be discerned in the end product" (*P&R* 42). Thus does Lakatos delicately distinguish and maintain the distinction between logical and historical analysis, in close parallel to the difference between the analytic capacities of Kantian Understanding, or *Verstand,* and the synthesizing capacity of Hegelian Reason, or *Vernunft.*

Pi does not directly answer Beta or Theta when they respectively suggest that a naive concept is replaced with "a more general, improved concept" or even a "totally different, novel concept" (*P&R* 90). Pi's answer is that exactly both these sides of the Hegelian Aufhebung occur immanently through the development of new language:

> Progress indeed replaces naive classification by theoretical classification, that is, by theory-generated (proof-generated, or if you like, explanation-generated) classification. Conjectures and concepts both have to pass through the purgatory of proofs and refutations. Naive conjectures and naive concepts are superseded by improved conjectures (theorems) and concepts (proof-generated or theoretical concepts) growing out of the method of proofs and refutations. And as theoretical ideas and concepts supersede naive ideas and concepts, theoretical language supersedes naive language. (*P&R* 91)

That is untendentiously a fine portrait of much of the outlook of the *Phenomenology* as well: that reason functions in time through language, with varying levels of sophistication attached to concepts and theories at different stages of growth, degeneration, and criticism. From this perspective, traditional logic is about, in Lakatos's terminology, "language-statics," while heuristic is about "language-dynamics" (*P&R* 93). Hegel, too, put great stock in the ability of a culture to articulate concepts in a rich vocabulary. For example, he admired the Greeks for the range of words they had for the sensible world, and considered language generally to constitute the medium of Bildung: "The language of a people is the best index of its *Bildung,* of its *Kultur* and of its *Aufklärung* [enlightenment] alike, and equally for its extent and depth."[18] In the *Phenomenology,* Hegel described changing social institutions, forms of thought, and conceptions of individuality; in *Proofs,* that morass of history is limited to theorems and their domains, and the languages used to describe them. Except for language change, the process Lakatos depicts is consistent with much of mathematical formalism; to include the mechanisms of language change in concept stretching and the development of theoretical language, as he does, is to identify what is essentially informal in the temporal development of a mathematical proof and what makes possible its genuine self-formation. While that historical development might be implicit in the mathematics of a given time, it is not "in" it but in mathematical history. With this proviso, Lakatos's method of proofs and refutations matches Hegel's phenomenological method almost exactly on the plane of language, concepts, and historically changing theories.

Riemann to Cantor

Before chapter 3 continues with the consequences Lakatos draws from his concept-stretching account of the method of proofs and refutations, it is useful to look briefly again at the "phylogenetic" counterpart in later-nineteenth-century mathematics. On the proof side, much of the improvements to rigor occurred by the time of Weierstrass, as noted earlier. As this happened, starting implicitly with Fourier and Dirichlet, and then with clear intent by Riemann, a second change in mathematical interest emerged about the object side of theorems: the introduction of functions and sets on the boundary of mathematical knowledge—not for their own sake but motivated by the expansionary growth described by Pi, using "pathological," monster-style functions to suggest new approaches to old problems. The exceptional sets of the exception-barrer in the latter half of the nineteenth century became of considerable theoretical interest. And just as analysis was finally becoming truly rigorous through Weierstrass's arithmeticization of proofs, only partially begun by Cauchy, the new objects being introduced into mathematics created another wave of false theorems, analogous to those earlier in the century, but now at a higher level of sophistication.

How, then, was Pi and Lakatos's notion of conceptual novelty, as indicated by changes in the objects of mathematical proofs, expressed in later-nineteenth-century mathematics, and how significant were these changes?

First, false proofs associated with many of the key concepts of analysis did persist throughout the century. It was common, for example, until explicit counterexamples were formulated by Riemann and Weierstrass, for texts to prove that continuous functions were differentiable. The false theorems of Cauchy's *Cours d'analyse* also persisted for years; for example, that the integral of a convergent series of functions may be integrated term by term and give the same value: If $\Sigma f_n(x) = F(x)$, then $\int F(x))dx = \int (\Sigma f_n(x))dx = \Sigma\int f_n(x)dx$. The interchange of summation and integration is valid for uniformly convergent series, but not pointwise ones—a type of error that occurred in several venues.[19] In contrast to the relative familiarity of Fourier series seen as counterexamples to Cauchy's false continuity theorem, counterexamples to unrestricted term-by-term integration were not obvious, nor were nondifferentiable continuous functions obvious, nor nonrectifiable curves not satisfying the classical formula for curve length, $\int [1 + f'(x)]^{1/2}\, dx$. The absence of counterexamples left the false theorems free from attack and were widely accepted as true. But while these many false theorems provide ample support for Lakatos's fallibilist account of nineteenth-century mathematics, and the method of proofs and refutations as the means by which they were later improved, they only offer modest examples of conceptual novelty. Cauchy's false theorem on term-by-term integration, for example, depends again on uniform convergence as the hidden condition, and distinguishing the convergence of a series of functions $f_n(x)$ from that of a series of terms x_n. Yet these innovations, significant as they were, are like those of Euler's theorem in *Proofs* in that they do not augur fundamental changes in mathematical concepts. Many improvements of Weierstrass over Cauchy in completing the arithmeticization of analysis were refinements of this kind, tending more toward "technical" improvements rather than radical novelty. Marginal objects, such as Dirichlet's characteristic function of the rationals, $\boldsymbol{Q}(x) = 1$ for x a rational number and 0 otherwise, raised questions of legitimacy, but the direction of response in the middle of the nineteenth century was often toward lemma-incorporating rigor rather than expanding conceptual scope. Where is the fundamental novelty Lakatos values so highly?

To see compelling cases of mathematical novelty requires looking not at a single theorem but a set of related theorems and their odysseys through the nineteenth century. This starts, again, with the philosophically central mathematical problem of that century: the representation of functions by Fourier or other trigonometric series. That is, given a function f and the infinite series defined by its Fourier coefficients, does the series converge to f and thus represent it? What conditions on f are needed to make that so? Seidel expressed the method of proofs and refutations using Dirichlet's 1829 proof of convergence of Fourier series for functions meeting what are called the Dirichlet conditions. These conditions include that f can *fail* to be continuous at a *finite* set of points—the "exceptional points" P of the function—and Dirichlet's convergence proof will still work. Dirichlet conjectured, but could not prove, as he wrote in a letter to Gauss published only in 1897, that his proof still worked for *infinite* discontinuity sets P, if P was what is today referred to as *nowhere dense:* for any $a < b$, there are $a < r < s < b$ such that on the interval (r,s), the function f is continuous and hence (r,s) is disjoint from the discontinuity set P: P is like a "dust." Dirichlet's unpublished lemma-incorporating insight is that his original lemmas do not require finiteness of P; they require only that the set of discontinuity points of f is nowhere dense. But in Dirichlet's time, and for almost the entire century, mathematical arguments about point sets such as infinite sets of discontinuity points of a function and their topology were completely unknown, including characterizations such as sets being dense, nowhere dense, countable, or uncountable. The real number continuum of rational and irrational numbers was, in this respect, largely uncharted territory, not even implicitly defined in terms of many "intended" properties because there was too little knowledge at the time to be implicit about. The conditions Lakatos discusses in *Proofs,* of a preexisting domain of polyhedra that was imperfectly picked out by false theorems, while possibly applying to types of convergent series, did not exist for point sets like P and the real number continuum even until midcentury: these mathematical objects were still yet to be born.

A major source for overcoming the challenges posed by Fourier series, and novel functions and sets, came from the young genius of Bernhard Riemann. He is known in calculus courses for his Riemann sums, an improvement over Cauchy's similar approach to partitioning the area under a graph, and then defining its integral through the limit of finer and finer partitions. Riemann's integrability conditions were far more sophisticated than Cauchy's and invited the difficult theoretical question of identifying the class of functions meeting them: what functions are Riemann integrable? At the same time, Riemann began the systematic use of pathological functions to identify improvements in theorems, such as nondifferentiable continuous functions, for which a published example appeared only in 1875. Today, stock and commodity prices graphed daily in the newspaper are commonly modeled using such functions. This explicit use of "nonnatural" functions, which Riemann even described as functions "not found in nature," became part of his heuristic to extend integration theory, with functions now allowed to go beyond either the old Euler-analytic or even physically interpretable functions. In this way, pathology became normalized through its productive use in Riemann's integration theory. A second area of Riemann's interest was the representation of functions by trigonometric series. Instead of extending Dirichlet's proof by reshaping the *sufficient* Dirichlet conditions, Riemann decided it was time to look at the *necessary* conditions for representation via a trigonometric series, not just Fourier series, and he proved a set of theorems establishing necessary and sufficient conditions for convergence to occur.

Nonetheless, the inchoate ideas of discontinuity point sets found in Dirichlet were not developed via Riemann in the context of trigonometric series but instead in integration theory. Riemann lived only from 1826 to 1866, and it was through the work of several of his successors that these two trajectories would intersect. It was a time, as put by Ivor Grattan-Guinness, of "Riemannian problems attacked by Weierstrassian methods."[20] With Riemann's new theories of functions and integration also came a series of false theorems characterizing classes of Riemann-integrable functions, just as earlier in the century there had been a series of false theorems on series convergence. Among these was a false theorem proved by Hermann Hankel, who perhaps more than anyone, was motivated by Riemann's new approach using functions without "general" or natural properties like continuity or integrability. Hankel attempted to develop a general theory to discriminate functions that were useful for the purposes of real analysis, and his approach was critical of both Euler-style analytic concepts and, at the other extreme, completely arbitrary functional correspondences. The former, Hankel pointed out, was not well defined, since there were analytic expressions even for the apparently "arbitrary" characteristic function of the rationals, $\boldsymbol{Q}(x)$.[21] Hankel's new function classes of continuous, pointwise discontinuous, and totally discontinuous functions, he believed, separated functions "amenable" to analysis from those outside its reaches. In particular, he proved a false theorem, which would persist in the literature for over a decade, that the Riemann-integrable functions were exactly his pointwise discontinuous functions.[22] The false assumption in Hankel's proof would be repeated by many mathematicians: that a Dirichlet-style nowhere dense set must have "zero content," meaning that the set can be covered by disjoint intervals whose total length can be made arbitrarily small. This error in confounding topology and measure was not even noticed initially by Georg Cantor, who would create much of modern point set theory. Only in England was Hankel's mistake seen in 1875 by H. J. S. Smith at Oxford, although it was not recognized on the Continent. In hindsight, the generic error was to assume that sets that are "small" in a topological sense will be so in a measure-theoretic sense: that is the hidden false lemma. Hankel's theorem was finally rejected as false in an 1878 textbook by the Italian Ulisse Dini, but without a counterexample, that is, a nowhere dense set of positive outer content. There simply were few examples of nowhere dense sets and little to keep even post-Weierstrassians from making such significant errors.

It is here hard to maintain that these problems in real analysis could be seen as problems of imperfect proofs over a largely stable domain of mathematical objects, even if they were poorly intended in the spirit of Lakatos's concept stretching. What those mathematical objects were or could be was very much in formation: the real number continuum was being rethought as a rich structure of novel point sets, and the notion of a function was expanding as well; still, there was as yet no theoretical apparatus to organize problems that were not even perceived, such as Hankel's need to distinguish nowhere dense sets and sets of positive outer content. For much of the nineteenth century, there was no clear characterization of the irrational numbers, much less the new sets needed for integration theory. The conditions of false theorems from the early part of the century were now being repeated by the very mathematicians who saw their role as completing the change to new standards of rigor, but now with concepts having no physical or mathematical precedents.

Principal pieces of the solution to the problems of point sets would come from Cantor's work on Fourier series, and would lead not just to a new understanding of the needs of Riemann's integration theory and analysis generally but to Cantor's entirely novel ideas for the mathematical infinite. Cantor's innovations came about not through integration theory but back in the other fundamental area studied by Riemann: the representation of functions by trigonometric series. A question left open by Riemann was the uniqueness of trigonometric representations for a function when they existed. If two trigonometric series converged to the same function *f*, did they have the same set of coefficients? After Riemann, Heinrich Heine in 1870 had proved uniqueness assuming the series converged to *f* uniformly except at a finite set of exceptional points where the series could diverge from the target function *f*. As with Dirichlet, exceptional sets played a key role in the conditions needed to make Heine's proof work, and he did not attempt a proof allowing infinite exceptional sets. But the need to deal with new types of point sets crossed over to the theory of trigonometric series from that of integration theory and the problems raised by Hankel's work.

In the 1870s, Georg Cantor took up the issue of the uniqueness of trigonometric representations, and especially that of weakening Heine's conditions. Through a series of papers, Cantor eliminated Heine's assumption of uniform convergence, and finally extended, as Dirichlet had partly seen almost a half century earlier, the categories of exceptional sets needed to improve the characterization of uniqueness conditions. Cantor's work began with the creative use of many of his predecessors' ideas in analysis, but his new path in addressing exceptional point sets, intended to improve a theorem of analysis, would lead to some of the most novel mathematical notions ever conceived. The first idea in improving Heine's result was to consider the transformation of a point set P to its "derived set" P', consisting of limit or accumulation points of P. For example, the limit point of the set $\{1 + 1/n \mid n \text{ a positive integer}\}$ is 1, and the limit points of the set of rational numbers is the set of all real numbers. If P is finite, then P' is the empty set, but P' could be any finite size, such as with $P = \{1/n\} \cup \{1 + 1/n\} \cup \{2 + 1/n\}$, for which $P' = \{0, 1, 2\}$. Or if P is $\{m + 1/n \mid m, n \text{ are positive integers}\}$, P' is the infinite set of positive integers, and P'' is the empty set. Thus, Cantor was led to his "first-species" sets for which $P^{(n)}$ is the empty set for some integer n, and he was able to establish the uniqueness of trigonometric series allowing for the failure of convergence on a first-species set P. This new characterization of point sets provided the needed context with which to clarify the ambiguities of integration theory, and Cantor began to sort out the varieties of sets confounded earlier: nowhere dense, first-species-type sets, and sets of outer measure zero. Cantor could build nowhere dense sets of any finite order, with first-species sets capturing the limited conception of nowhere dense sets held by Dirichlet, Hankel, and Rudolf Lipschitz, who before Cantor had partially extended Dirichlet's work. To prove his uniqueness theorem, Cantor limited himself to the first-species sets, although he quickly saw the possibility of derived sets of higher order, for which *no* $P^{(n)}$ is finite. As such, one may define the intersection of the $P^{(n)}$ to be a new set $P^{(\infty)}$, and iterate the process of taking derived sets again:

$$P^{(\infty)}, P^{(\infty)\prime}, P^{(\infty)\prime\prime}, \ldots P^{(\infty)\infty}. \ldots$$

It was sequences of derived sets such as these that led directly to Cantor's idea of transfinite ordinal numbers, surely a novel mathematical concept if there ever was one. What Cantor saw was one of Lakatos's proof-generated concepts, the new underlying sequence of infinite ordinal numbers indexing the derived sets:

$$0, 1, 2, \ldots \omega, \omega + 1, \omega + 2, \ldots \omega + \omega. \ldots$$

Here, ω is the first transfinite ordinal number. The abstraction of the series of transfinite ordinals from the construction of derived sets was almost automatic, arising as Cantor wrote in retrospect, as a "dialectical generation of concepts."[23] Cantor formulated these ideas about higher-order derived sets early on during his study of trigonometric convergence, not later, when he moved on to set theory generally. The concepts of a limit point, derived sets, and then transfinite sequences of derived sets all arose through Cantor's goal of improving his theorems on trigonometric series, out of which he developed the first rigorous theory of transfinite numbers, and that soon took on an important life of their own. It is these several related notions from analysis that are the basic elements laying the foundation for Cantor's set theory. The improvement of analysis, and the theory of trigonometric series in particular, constituted the step, or steps, leading directly to the transfinite ordinals ω, $\omega + 1$, $\omega + 2$, . . . , and then later to the infinite cardinal numbers, or alephs, $\aleph_0$, $\aleph_1$, . . . $\aleph_\omega$, $\aleph_\omega + 1$, . . . and their relation to $2^{\aleph_0}$, the cardinality of the real number continuum. Cantor's continuum hypothesis, that $2^{\aleph_0} = \aleph_1$, was shown to be independent of the Zermelo-Fraenkel axioms for set theory only in 1963.

These cardinalities and the sets associated with them are among the most remarkable creations in the history of mathematics, extending far beyond any possible implicit notions of the real number line. To be sure, much else was needed to create Cantorian set theory. Some of these ideas still were motivated by problems in real analysis that required clarification of different conceptions of size. Nonetheless, Cantor's infinities and their affiliated new objects stem directly from his initial approach to improving a classical theorem of analysis via his derived sets and the invention of the sequence of transfinite ordinal numbers. Perhaps the $\aleph$s are platonically real in some substantive sense and Cantor just discovered them; who knows? But present-day *knowledge* of them is tied to a history whose main features, especially in terms of the emergence of new mathematical concepts, is characterized in its main lines by the method of proofs and refutations. The weird sets and functions studied by Cantor were pathological, but they served to extend mathematics in useful ways, even continuing Riemann and Hankel's programs. Henri Poincaré appears occasionally in *Proofs* as a caricatured opponent of the recognized age of pathological functions in late-nineteenth-century mathematics, although from this perspective, his criticisms would not be supported by ideas of creating new mathematical objects in *Proofs,* nor the origins of modern Cantorian set theory. The second half of the nineteenth century provides as rich a study in the erroneous yet ultimately beneficial formulation of emerging functions and point sets as did the first half century for unrigorous proofs. The creation of mathematically novel objects reaches an apotheosis with Cantor's infinities, and methodologically much of this novelty may be attributed to the historical problem of improving the theorems on representing functions by trigonometric series. As Lakatos would say, Cantor—following Heine, Riemann, Dirichlet, and others—did not prove what he set out to prove, and mathematics is the greater for it. As characterized in *Proofs,* the expression of mathematical novelty in mathematical objects, so crucial to the trajectory of late-nineteenth- and early-twentieth-century mathematics, is intimately related to the criticism and improvement of theorems through the method of proofs and refutations.

The leapfrogging of theorem improvement and articulation of new mathematical entities continued during the twentieth century, even for the relatively narrow problems of real analysis. To successfully extend the concept of the integral, and to make use of a point set perspective in real analysis generally, required both Henri Lebesgue's measure theory and Émile Borel's hierarchy of point sets, a large domain over which Lebesgue integration is valid. Was Riemann's quest to characterize integrable functions, or now sets, finally answered? Had an intended domain finally been pinned down? No, not unless one is willing to make special assumptions about the kind of set theory used as the background for one's integration theory. Some set theories imply that *all* sets of real numbers are Lebesgue measurable, while others imply the existence of nonmeasurable sets. Hankel's problem of settling on a boundary for functions or sets "useful" for analysis kept moving, and as was discovered in the twentieth century, the question of where that boundary can lie can itself be formally treated through mathematical independence proofs; it is no longer a contentious matter of monsters inside or outside mathematics.[24] From the perspective of *Proofs,* it would appear to be a pseudoquestion without answer, though, as to which one of those formal options is correct in some primal, Platonic, or other ontological sense. The movement of mathematical thinking outside its boundaries and the creation of mathematical novelty, according to *Proofs,* are natural consequences of the methodology of proof embraced by nineteenth-century mathematicians, not the discovery of a preexisting realm.

When, for example, Ernst Zermelo was criticized for introducing general choice functions, he stimulated a major controversy by going beyond, it was thought, even the Dirichlet concept of function as arbitrary correspondence. In France, René Baire, Borel, and Lebesgue all voiced considerable opposition to Zermelo's ideas as transgressing the boundaries of mathematics, but perhaps also as a means of defending measure theory, since nonmeasurable sets, among many other second-generation pathological monsters, can be constructed using choice functions, and not constructed without them. Whether choice functions are "real," therefore, was an important way of charting the future of integration theory, just as earlier generations had grappled with other pathological functions. Jacques Hadamard defended Zermelo against

Lebesgue's demand for a law defining choice functions, like a Eulerian analytic expression, as a necessary shift in the boundaries of mathematical concepts: "The essential progress of mathematics, since the invention of the infinitesimal, has consisted in the annexation of successive notions which were, some for the Greeks, others for the geometers of the Renaissance or the predecessors of Riemann, 'outside of mathematics' because they were impossible to describe."[25] Lakatos makes this antithesis of exception- and monster-barring a central feature of lemma-incorporation and the creation of new mathematical concepts. In this way, the changes brought about in mathematical objects over the last two hundred years would appear to depend significantly on the theorem-improving technique of the method of proofs and refutations.

Appendix: Some Mathematical Background

First, some elements of real analysis, or the study of the real number line, will be summarized. A central concept of mathematics is that of a function. A function $f(x)$ can be given by a rule, such as $f(x) = x^2 + 1$. That means that to calculate $f(x)$ for, say, $x = 1$, $1^2 + 1 = 2$ is calculated. Similarly, $f(0) = 1$, $f(-1) = 2$, and $f(\sqrt{2}) = 3$. Sometimes one also writes $y = f(x)$, and the graph of a function includes all the (x, y) pairs representing $f(x)$ on a coordinate grid. The function $x^2 + 1$ is said to be continuous, roughly because the graph can be drawn without having to raise a pencil at any point, and the graph does not swing about or oscillate too wildly (see figure 3). The function $f(x) = x^2 + 1$ is also differentiable, meaning that the graph is smooth or has no sharp corners. The function $g(x) = (\sqrt{x^2})$, which gives the absolute value of x, is continuous, but not differentiable because of a "corner" at $x = 0$, though $g(x)$ is differentiable everywhere else. Both $f(x)$ and $g(x)$ are defined by simple equations, and for much of the nineteenth century, providing such an analytic expression was the only legitimate means for defining a function. Today, however, many more functions are allowed into mathematics. A function can be any "arbitrary" correspondence, not just one specified by a rule. For example, the graph of a stock price across an entire year is fairly arbitrary, with no single rule describing the up-and-down patterns. A function today is thought of as any correspondence, formally a set of ordered pairs $\langle x,y \rangle$ in which only a single value y corresponds to any given x in the function's domain. Hence, the answer to, "What are allowable functions?" depends on answering, "What is an allowable set?"

The idea of an arbitrary functional correspondence was suggested by Leonhard Euler around 1750, but it was controversial until at least after the mid-

1800s. Much of the controversy arose around infinite series and series of functions. An example of a series is $1/2 + 1/4 + 1/8 + \ldots$, written $\Sigma 1/2^n$, with n ranging over 1, 2, 3, . . . Imagine a square whose sides are 1 unit long. Divide the square in half. Half the entire area is 1/2, the first term in the series. Now divide one of the halves in two, so that each subarea is 1/4, the second term in the series. Keep subdividing, and each smaller area is $1/2^n$, which means that the infinite series $\Sigma 1/2^n$ must sum to the area of the entire square, or 1. That is not a rigorous proof but it illustrates what is meant by a convergent series, or one that sums to a finite value. Other series diverge, such as $1 + 1 + 1 \ldots$, which just keeps getting bigger as more terms are taken.

A series of functions can also be created, like $f(x) = \Sigma x^n$, with $n = 1, 2, \ldots$ In this case, $f(0) = 0$, $f(1/2) = 1$, and for $x = 1$ the series diverges, as it does for any $x > 1$. If we write $f_n(x) = x^n$, then $f(x)$ is the sum $\Sigma f_n(x)$. The usefulness of series of functions is that the f_n's may be easy to calculate, and a partial sum may provide a good approximation for $f(x)$. Many computer algorithms today utilize just such series to calculate different functions, using sufficiently many terms to get an adequate approximation to the value of the entire infinite sum. The sine function from trigonometry, for example, can be calculated using the series $\sin(x) = x - x^3/3! + x^5/5! - \ldots$, where $n!$ means $n \times (n-1) \times \ldots \times 1$. Similarly, the cosine function $\cos(x) = 1 - x^2/2! + x^4/4! - \ldots$. Cos(0) can be calculated, for instance, as $\cos(0) = 1 - 0^2/2! + 0^4/4! - \ldots = 1$.

These several ideas—of what a function is, convergent and divergent series, continuity and discontinuity—became embroiled in one of the great controversies of nineteenth-century mathematics: the discovery of what are now called Fourier series, named for the French mathematician Joseph Fourier. A Fourier series can be created starting with many functions $f(x)$, and expresses $f(x)$ as the infinite sum of trigonometric sine and cosine functions. This series for a function $f(x)$ thus has the form $f(x) = \Sigma\, a_n \sin(nx) + b_n \cos(nx)$, with the a's and b's, whose definitions shall be skipped here, are numbers called the Fourier coefficients. The idea is that a typical curve on an interval could be decomposed into an infinite sum of cyclical sine and cosine functions, like decomposing the harmonics of an arbitrary sound. Fourier's result was counterintuitive because he was *not* assuming that his original "sound" was necessarily regular, like a musical note, or a nicely vibrating string; it could be pretty much any curve at all. Imagine, for example, a heated slab, whose temperature at any point was simply "given," not specified by a particular rule. A cold slab could even be joined to a warm slab at their edges, creating a discontinuous "jump" in temperature. Fourier claimed that his theorem showed how such an arbitrary temperature function could be expressed as a sum of sines and cosines. Fourier's proof was notoriously unrigorous, and would not be repaired until 1829 by Dirichlet. But his series seemed to work for many functions, and Fourier was awarded a prize for his work in 1812.

Here is a surprising instance of such a series provided by Fourier in his 1807 memoir submitted to the Academy of Science of Paris: $\sin(x) - 1/2 \sin(2x) + 1/3 \sin(3x) - 1/4 \sin(4x) + \ldots = \Sigma (-1)^{n+1} \sin(nx) / n$. The graph of this series is shown in figure 4. As can be seen, the limiting function is discontinuous at the points $\ldots -3\pi, -\pi, \pi, 3\pi, \ldots$ even though the series converges for all x, and is composed of a set of ordinary continuous functions. Hence, to summarize a much more complicated history, Fourier series simultaneously raised two significant problems. First, what *were* functions? Could they be "arbitrary" like the temperature of a hot slab, or the arbitrary positioning of a string, and not defined by a clear rule? Second, what was happening with the convergence of series? How could a set of continuous functions converge to a discontinuous function? This latter problem, of continuity and convergence, turned out to depend on a lack of clarity about the type of convergence possible for series. A series could converge for every point x, or pointwise, but the accuracy of the convergence to the limiting value might or might not be uniform for all values x. The infinite series for $\sin(x)$ and $\cos(x)$ above turn out to be uniformly convergent, but Fourier's example just given is only pointwise convergent. The distinction of pointwise and uniform convergence, and the validity of Fourier series, took several decades to clear up. The question of mathematically allowable functions would persist through the nineteenth century, and is still with us today.

The integral $\int_a^b f(x)\mathrm{d}x$ of a function $f(x)$ can be thought of as the area under its graph between the points a and b, except that if the function is negative, the area is taken as negative also. The dx notation is meant to suggest a "small" length of x_0 at the base of a rectangle R with height $f(x_0)$, and whose area would be $x_0 \times f(x_0)$. The $\int$ is intended to suggest an infinite continuous sum, so that even if the graph of $f(x)$ is curved, making the area of rectangle R an approximation, the area under the curve can be calculated as a limit of increasingly narrow rectangles.

The *measure* of a set generalizes that of points, lines, areas, and volumes. A point has measure zero, but the measure of a straight line is its length, and the measure of a plot of land is its ordinary area. To generalize integrals $\int_a^b f(x)dx$ to functions not defined by rules, like Dirichlet's $\boldsymbol{Q}(x)$, discussed earlier, a conception of measure is needed that allows making sense of the "area" under the "graph" of $\boldsymbol{Q}(x)$. These measures are today used even in

Curve *(a)* is defined by the rule $f(x) = x^2$. It is continuous and differentiable, or "smooth." Curve *(b)* is continuous but nondifferentiable at its corners, and *(c)* is discontinuous at the breaks. For some nineteenth-century mathematicians, *(b)* and *(c)* were not single functions but multiple ones pasted together, each "piece" being defined by a different analytic rule.

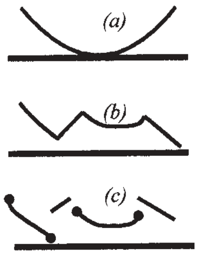
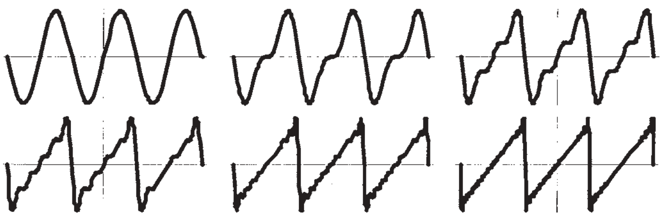


Figure 3. Types of functions



The Fourier sine series $\sin(x) - {}^1/_2 \sin(2x) + {}^1/_3 \sin(3x) - \ldots$ converges to the sawtooth function below. At odd multiples of π, the series converges to zero but is discontinuous. Shown above is the series taking 1, 2, 3, 5, 10, and 20 terms.

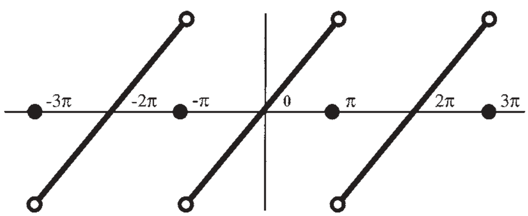


Figure 4. Trigonometric series example.

applied mathematics, such as probability theory, where events get weighted, or measured, by their likelihood of occurrence, such as the chance of a stock price exceeding a given value. A stock price, or the motion of a dust mote suspended in water, are instances of continuous but nondifferentiable functions, and it is not obvious how the "rectangles" that measure the area under their graphs are created. Once twentieth-century mathematicians decided they needed more general concepts of measure and integration—for example, through the useful Lebesgue integral—they were again led to controversies regarding the functions allowable in mathematics, as summarized in the previous section of this chapter.

The following concepts are from logic and the foundations of mathematics.

Peano Arithmetic refers to a set of axioms for the natural or counting numbers, 0, 1, 2, 3 . . . , intended to capture their basic properties. Roughly, the Peano axioms include the definitions for the numbers 0 and 1, addition, and multiplication, plus the rule of mathematical induction: if a property P is true of the number 0, and the truth of P for the number n implies its truth for $n + 1$, then the property must be true for all numbers. For example, if P says "the sum of the first n integers is $n \times (n + 1) / 2$," $P(0) = 0$, so the inductive base is true. As such, assuming the so-called inductive hypothesis, that $1 + 2 + \ldots + n = n \times (n + 1) / 2$, if $(n + 1)$ is added to both sides of the equation, then $1 + 2 + \ldots + n + (n + 1) = n \times (n + 1) / 2 + (n + 1) = (n^2 + n) / 2 + (n + 1) = (n + 1) \times (n + 2) / 2$, which is just the "$(n + 1)$" version of the formula. Hence, by mathematical induction, the formula is true for all numbers n. The axioms, or variants of them, are named for the nineteenth-century Italian mathematician Giuseppe Peano, who devised one of the earliest systems of modern symbolic notation.

Zermelo-Fraenkel set theory, or Zermelo-Fraenkel-Skolem set theory, or ZF, refers to a group of axioms about set membership, denoted "ϵ," with the intended meaning of "$x \epsilon A$" being "x is a member of set A," with x itself being a set, too. The axioms, very roughly, state that sets "exist," for instance, that an "empty set" $\varnothing$ exists, and that other sets can be formed, like the set consisting of the empty set, or $\{\varnothing\}$, and then $\{\{\varnothing\}\}$, and so on; or from sets x and y, their union $x \cup y$ also exists. An additional axiom that an infinite set exists ensures that that process can go on long enough to define sets that can stand in for the natural numbers. That is, we can define the natural numbers $N = 0, 1, 2, \ldots$ using the sets $\varnothing$, $\{\varnothing\}$, $\{\{\varnothing\}\}$, . . . The axiom of infinity guarantees the existence of a set containing all of these at once. Another axiom states that the power set, $\wp(x)$, or set of all subsets of a given set x, is again a set, and other axioms allow building up other sets needed for mathematics. The real number line, including all the rational and irrational numbers, is usually identified with the power set $\wp(\boldsymbol{N})$ of the set of natural numbers $\boldsymbol{N}$; basically, one thinks of a number as expressed in binary, or 0–1 notation, and then a subset of $\boldsymbol{N}$ can be defined by the places at which the 1's are used.

ZF can be extended with additional axioms, such as the axiom of choice, or AC, which says that if $\{x_a\}$ is a set of sets, then a "choice function" f exists with the property that $f(x_a) \in x_a$. Choice functions sound trivial, but they turn out to be essential for certain parts of mathematics. Surprisingly, they cannot be proven to exist in general using ZF alone: they must be assumed as an additional axiom. Bertrand Russell likened a choice function to a means of choosing one sock each from an infinite set of pairs of socks, but without the socks being characterized as right versus left, or brown versus black: mathematically, the ability to choose one from each pair is guaranteed only by the axiom of choice. Kurt Gödel proved the consistency of the axiom of choice with ZF in 1938, and Paul Cohen showed its independence in 1963. Many philosophers and mathematicians think of ZF plus the axiom of choice (ZFC), or one of their variants as providing a foundation for mathematics because, once numbers, functions, and other mathematical objects are all identified with particular sets, almost all of mathematics can be seen to be derivable in, say, ZFC. Something of a countermovement is now forming against the long-standing ZFC orthodoxy in favor of much weaker systems that suffice for most useful mathematics.[26]

The first infinite ordinal number ω can be thought of as the first number "beyond" the sequence 1, 2, 3, . . . Then, the next ordinal number is its "successor," or $\omega + 1$, and the sequence continues $\omega + 1$, $\omega + 2$, $\omega + 3$, . . . The next ordinal after all those is then $\omega + \omega$, and so on. Infinite cardinal numbers are used to characterize different sizes of infinity. The least infinite cardinal is the size of the set of natural numbers 1, 2, 3, . . . , and is denoted $\aleph_0$, or "aleph-zero." The next cardinal number is $\aleph_1$, and the sequence of infinite ordinal numbers allows for an infinite continuation of the cardinal numbers: $\aleph_0, \aleph_1, \ldots \aleph_\omega, \aleph_{\omega+1}, \ldots$ The initial transfinite ordinals including ω, $\omega + 1$, $\omega + 2$, . . . ,$\omega + \omega$, and much further, all turn out to have the same cardinal size, namely $\aleph_0$. The ordinals are primarily an "index" set used to "enumerate" the $\aleph$'s and lots of other sets, yet some infinite ordinals are also infinite cardinals. The cardinal $\aleph$'s are ordinals as well, just "big enough" to be "larger" than their predecessors. Set theorists think of the universe of sets as having a backbone of ordinals, with all sets appearing at various ordinal levels. Georg Cantor initiated the study of infinite ordinal and cardinal numbers during the late nineteenth century.

Cantor's continuum problem is to determine which one of the $\aleph$'s corresponds to the size of the ordinary real number line, or even just the size of the set of real numbers between 0 and 1. Cantor's original continuum hypothesis was that the cardinality of the real numbers was the second infinite cardinal, $\aleph_1$. It turns out that neither ZF nor ZFC, nor many other reasonable set theories, determines the answer. It can be assumed that the cardinality of the set of real numbers can be just about any of the $\aleph$'s desired. Cantor's continuum hypothesis, then, is seriously undecidable with respect to ZF and many set theories. One response is the need to just discover or invent the right axioms to add to ZF in order to decide the continuum hypothesis, such as axioms of higher infinities, beyond those guaranteed by ZF. Another response is that the cardinality of the real numbers is indeterminate, and the quest for resolving it is a will-o'-the-wisp.

First-order logic is the most common formal approach for representing mathematical theories within mathematics. Very roughly and imprecisely: A first-order language is built up using symbols $\forall$, $\exists$, $\rightarrow$, $\sim$, (,), =, a set of variables $x, y, z, \ldots$, and predicate symbols $P, Q, R, \ldots$ Often, symbols such as $\in$, +, $\times$, $<$, and so on are used instead of letters, but these purely syntactic symbols should be distinguished from the mathematical entities to which they can refer. Elementary formulas are, for example, $x = y$, $x = x$, and $P(x,y)$, assuming P is a two-place predicate, say, intended to represent "*a* is less than *b*." Or $P(x,y)$ could represent "*b* is the next counting number after *a*." A functional relation like addition, $a + b = c$, can be represented using a three-place $Q(x, y, z)$. More general formulas are defined inductively: If ϕ and ψ are formulas, then so are $(\forall x)(\phi)$, $(\exists x)(\phi)$, $\phi \;\&\; \psi$, $\sim\phi$, intended respectively to represent "for all x, ϕ", "for some x, ϕ", "ϕ and ψ", and "not ϕ". Other connectives, such as $\rightarrow$ for "implies," can be defined using other primitives—for example, $P \rightarrow Q$ can be defined as $\sim(P \;\& \sim Q)$. Formalized sentences represent mathematical statements like "there exists an infinite number of prime numbers," or "the sum of two odd numbers is an even number," or "a quadratic equation has at most two solutions." A proof is a special syntactic relationship between formulas or sentences, in which a formula or sentence ϕ is said to be derivable from $\phi_1, \ldots \phi_n$. A first-order theory is a consistent, finite, or infinite set of sentences. Peano Arithmetic, ZF, ZFC, and so on are examples of such formalized theories. Consistency means that no contradiction, such as $(P \;\& \sim P)$, which is always false, can be deduced from the theory.

Many first-order theories exist for the arithmetic of addition and multiplication, like Peano Arithmetic, and include axioms for addition and multiplication, such as the formalization of "*a* times (*b* plus *c*) equals *a* times *b* plus *a* times *c*". When these theories contain just a minimum of assumptions for the properties of addition and multiplication, then the syntactic or formal notion of "the sequence of sentences $\phi_1, \ldots \phi_n$ is a proof of ϕ" can itself be "coded" as a sentence about numbers. In ZF or other set theories, there is the additional step of coding numbers as sets, but the procedure is the same. The idea is to use the prime factorization of a number to track different sequences, and manipulate statements about numbers and their factorizations as statements about proofs. Such internal representations of mathematical proof *within* a formalized theory is often called metamathematical. The technique of representing formulas, sentences, theories, and proofs via number coding is called Gödel numbering, after Kurt Gödel, who first devised the method. Gödel showed, in 1931, that many consistent mathematical theories *T* including just a minimum of axioms for addition and multiplication are incomplete, meaning that there is a sentence such that neither it nor its negation is provable in *T*. Gödel's work also implied that this incompleteness could not really be eliminated from much of mathematics. Moreover, Gödel showed that many theories strong enough for some simple mathematics cannot prove the metamathematical sentence asserting the consistency of *T*; these theories, in other words, cannot prove their own consistency. Gödel's theorems therefore provide mathematically rigorous, skeptical responses to proposals for creating certain foundations for mathematics; one example is the Hilbert program intended to establish completeness and consistency for infinitary mathematics on a finitary basis. Gödel's proofs and some of their history are discussed in chapter 4.

This page intentionally left blank

3 | Mathematical Skepticism

The significance of lemma-incorporation for contemporary criteria of mathematical truth becomes explicit as chapter 1 of *Proofs* draws to a close. "The most important developments in logical theory," writes Lakatos, "were usually preceded by the development of mathematical criticism" (*P&R* 105n1); similarly, "The concepts of criticism, counterexample, consequence, truth, and proof are inseparable; when they change, the primary change occurs in the concept of criticism and changes in the others follow" (*P&R* 104); and the sea change that took place after 1900 was not merely a "turning point in the history of mathematics" but a "revolution in mathematical criticism [which] changed the concept of mathematical truth, changed the standards of mathematical proof, changed the patterns of mathematical growth" (ibid.). The claim may be overstated, yet Lakatos evidently believes the method of proofs and refutations is a kind of conceptual ur-process underlying twentieth-century logical theory.

What Lakatos wants his historical reconstruction to imply for the logical theories of this century is that mathematical progress implies no natural boundary defining the class of terms to which concept stretching may apply, and that even the logical terms of a theorem may be stretched in order to provide a refutation. Concept stretching and lemma-incorporation are the mechanisms of progress, but

> for any proposition there is always some sufficiently narrow interpretation of its terms, such that it turns out true, and some sufficiently wide interpretation such that it turns out false. Which interpretation is intended and which unintended depends of course on our intentions. . . . But concept-stretching will refute any statement, and will leave no true statement whatsoever. (*P&R* 99)

That is a remarkable assertion about the flexibility of informal language, comparable to Hegel's claim that "any so-called basic proposition or principle of philosophy, if true, is also false simply insofar as it is merely a basic proposition or principle. It is therefore easy to refute it. The refutation consists in demonstrating its deficiency" (*Pf* 36).

In *Proofs,* the previous skeptical proclamation is provided by Kappa, who delivers several surprising examples of concept stretching in the history of Euler's theorem to make the point that you cannot both have certainty and meaning in mathematics. Mathematics, for Kappa, as far as it is meaningful, is subject to concept stretching because it is part of informal, natural language, and only "gibberish is safe from refutations" (*P&R* 102), meaning the uninterpreted signs of a mechanical calculus or syntax. The preparation for this skeptical confrontation has been laid throughout *Proofs.* When Teacher is attacked early on for disbelieving that a list of exceptions to Euler's theorem is exhaustive, he guardedly responds that "I do not even need a concrete new 'exception' for my argument. My argument was for the *possibility* of further exceptions" (*P&R* 27). When facing Beta-Cauchy, the exception-barrer who thinks all monsters have been excluded, who asks, "Could you offer a counterexample?" Teacher replies, "You cannot know that I shall not. You *improved* the original conjecture, but you cannot claim to have *perfected* the conjecture, to have achieved perfect rigor in your proof" (*P&R* 30). Then Kappa, in a discussion with the Euclidean Epsilon, asserts that "there is an infinite regress in proofs; therefore proofs do not prove" (*P&R* 40), and Lakatos in a footnote remarks that "knowledge has no foundations" (*P&R* 54n2). Or Alpha, again leveraging the imprecision of natural language, observes: "Since the vagueness of language makes the *rigor of proof-analysis* unattainable, and turns theorem-formation into an unending process, why bother about the theorem?" (*P&R* 51). Then Lakatos again: "Few mathematicians would confess—like Beta—that reason is too weak to justify itself" (*P&R* 54n2). Thus is Lakatos's mathematical fallibilism expressed in *Proofs* through repeated skeptical rejoinders to attempted claims to certainty—either that of a final, unassailable proof, or an exhaustive exceptions list, or a priori true axioms, or sure intuitions.

All these confrontations between the shapes of dogmatism and their skeptical rejoinders end with the central dogmatist-skeptical confrontation of *Proofs:* reconciling the development of nineteenth-century mathematical content and rigor. The method of proofs and refutations increases both, but for Lakatos, these are two values participating in a fundamental trade-off: rigor only becomes absolute by giving up the ability to increase content through concept stretching. There are, Lakatos argues via a continued examination of Euler's proof, no absolutely "inelastic, exact concepts" in informal mathematics, and there is no a priori boundary between rational and irrational concept stretching, in which one may "stretch any term you like just as you like" (*P&R* 102). Concept stretching has no natural end point that closes down possibilities for unexpected refutation or theorem reformulation: it is just not known where in the body of mathematics we either might turn out to be wrong or choose to be wrong.[1] The variety of mutually inconsistent foundations for mathematics now known shows that such attacks need not depend on a trivial reinterpretation of logical or set- or number-theoretical concepts. L. E. J. Brouwer's mathematical intuitionism, to cite one case, is a guidebook to constructing examples of this kind regardless of what one wishes to stretch: the equality relation, operations on real numbers, logical constants, and methods of inference. For instance, in the nonstandard interpretation of logical terms in intuitionism, even the classical law of noncontradiction is no longer a logical truth. The example is unorthodox, although it illustrates Lakatos's point that concept stretching within informal mathematics is always possible in principle: that is the skeptical threat.

One may object to Lakatos's claim that counterexamples can always be produced by concept stretching, as Lakatos's editors, John Worrall and Elie Zahar, do when they say that "by definition, a valid proof is one in which, no matter how one interprets the descriptive terms, one never produces a counterexample, i.e., its validity does not depend on the meaning of the descriptive terms, which can thus be stretched however one likes" (*P&R* 100n1*). But the proofs they refer to are formal ones—mathematical objects that may or may not capture all the features of the informal proofs that are the topic of *Proofs.* In particular, their proofs must be in a fixed language, unless they have one that already accommodates all the linguistic changes and reinterpretations possible through concept stretching. For Lakatos, the relation between informal proofs and their formal representatives is that the representation of a theorem by a formal proof requires an appropriate translation into the formalism. As he says, "The validity of an intuitive inference depends also on the translation of the inference from ordinary (or arithmetical, geometrical, etc.) language into the logical language: it depends on the translation we adopt" (*P&R* 123). *After* translation, concept stretching does not occur by fiat, just as one settles on a definition of "edge," "vertex," or "face." But before translating, or after recognizing the defects of a given formalization, it is possible in principle to analyze any theorem and stretch even those logical constants that cannot be stretched within a formal proof. Hence, the options for what counts as a translation also condition and limit the notion of logical validity.

A certain foundation for mathematical knowledge demands, says Lakatos, a final, "crystal-clear definition of 'counterexample,'" which in turn, "amounts to a demand for crystal-clear, inelastic concepts in the metalanguage as a condition for rational discussion" (*P&R* 102n2). If there are no such concepts, though, or even if one only skeptically believes there might be no such concepts, every standard of informal mathematical truth can be potentially refuted. This again is a form of Lakatos's assertion that any informal mathematical proposition can be reinterpreted as true or false. The possibility occurs, then, that "unlimited concept-stretching destroys meaning and truth" (*P&R* 99). Meaning is destroyed if we do not settle on a provisional, though marginally vague set of primitives, on which a standard may rest. Truth is destroyed if no statement is distinguished from another by a stable truth-value as every conjecture is falsified: a fallibilist mathematics on fast-forward and out of control. To make the same point more constructively, while modern mathematics for Lakatos cannot do without concept stretching for its growth, and while its coherence is maintained by delimiting boundaries that prevent unlimited concept-stretching refutations, the price of certain kinds of progress is the limited historical affirmation of unbridled mathematical skepticism.

In terms of the mathematical history of *Proofs,* Lakatos implies that the foundational crises in mathematics at the turn of the nineteenth century were, in part, the historical realization of skeptical antifoundationalism implicit in the method of proofs and refutations. There was, on various fronts, as full a suspension of criteria of mathematical truth as one is likely ever actually to see, and these changes, brought about by deploying the power of mathematical concept stretching against logical problems, complete the historical odyssey of *Proofs.* The method of proofs and refutations is inherently a skeptical powerhouse, and as logical problems were translated into mathematical ones, the solutions became subject to criticism without definitive end. The collapse into a skeptical abyss was tempered by the development of modern logical theory, and the provision of a relatively safe, if flexible, framework in which foundational issues could be localized and investigated without threatening all of mathematics. This process, in which we still participate, was the creation of modern theories of logical form and the concept-stretching transformation of "mathematical truth into logical truth" (*P&R* 102). The stretched concepts are the usual ones of logical theory: truth, proof, validity, and inference, and the semantic and syntactic relations among them. This view of formal logic does not depend on any prior characterization of logical and nonlogical terms or concepts; it relies only on whether some, or all, concepts are simultaneously subjected to concept-stretching refutation.

In *Proofs,* this is the main historical change occurring after the instillation of Weierstrassian rigor and proof analysis. In the reconstructed dialogue of *Proofs,* the historical demarcation between open, radical concept stretching and mitigated, constructive concept stretching is marked by Theta's proposal of Bernard Bolzano's untimely, because modern, 1837 definition of relativized logical truth. In Lakatos's framework, Bolzano provides the means to acknowledge the possibility of Kappa's skeptical abyss without rejecting the basis of skeptically driven, lemma-incorporating refutation. Wholesale chaos is prevented by selecting out some "logical constants," which leaves the remaining typically number- or set-theoretical "constituents" open to stretching and criticism. Logical truth will then not depend on the meaning of these stretchable terms. What happened historically, says Lakatos, is that

> nineteenth-century mathematical criticism stretched more and more concepts, and shifted the meaning-load of more and more terms onto the logical form of the propositions and onto the meaning of the few (as yet) unstretched terms. In the 1930s this process seemed to slow down and the demarcation line between unstretchable ('logical') terms and stretchable ('descriptive') terms seemed to become stable. A list, containing a small number of logical terms, came to be widely agreed upon, so that a general definition of logical truth became possible; logical truth was no longer 'with respect to' an ad hoc list of constituents. [Alfred] Tarski [in 1935] was, however, puzzled about this demarcation and wondered whether, after all, he would have to return to a relativized concept of counterexample, and consequently, of logical truth—like Bolzano's. (*P&R* 103n2)

From Lakatos's perspective, we never really left Bolzano's framework. As Lakatos wrote elsewhere, this process eventuated in "a sort of non-logical mathematical necessity" (*MSE* 127).[2] At different stages in the development of modern foundational studies, different standards—such as for logical connectives, quantifiers, implication, and so on—allowing different Bolzano-like "sets of constituents," and what Lakatos elsewhere calls "quasi-logical constants" (ibid.). The development ending with Tarski just noted may be rephrased as "the gradual elimination of quasi-logical constants" and is "one of the most essential features of the history of mathematical rigor" (ibid.). But since the characterization of rigor in *Proofs* also depicts mathematical truth, this creative historical process explains why the main message of

Proofs is, as Lakatos indicates, that "criticism does not assume a fully articulated deductive structure, it creates it" (*MSRP* 45n4). The mathematical criticism of the nineteenth century, in this view, created what is recognized as mathematical deductive structure today. Hence, for the mathematical present, in Lakatos's account, underlying the variety of contemporary foundations is an essentially skeptical core, stabilized through present-day knowledge of alternative set-theoretic and logical assumptions.

At this point, the reader of *Proofs* may be disappointed with the generality of this account of formalism and logical theory. It lacks the historical definition and precision of the story of lemma-incorporation. For example, Lakatos largely vaults over the entire era of pathological functions that was both the methodological bridge in analysis between the "high" proof analysis of Weierstrass and logical theory, and the birthplace of the fundamental ideas of Cantorian set theory.[3] It is also difficult to see just how Lakatos's mathematical skepticism may be embedded in specific problems of contemporary mathematics. If, as Lakatos's teacher Árpád Szabó says, "the whole point of studying the history of mathematics is to try to understand how the development of science has led from one problem to another," the skeptical conditions behind lemma-incorporation should show themselves in a more mathematically definite way.[4] Chapter 4 will fill this gap of better characterizing the mathematical present as history with a final piece of heuristic analysis uniting two important themes of *Proofs:* the inherent skepticism of the method of proofs and refutations, and the idea that formalisms are themselves part of informal mathematics translated out of the latter.

To do so, it is necessary to return one last time to Hegel's phenomenological method. As mentioned above, there is a close homology, if not identity, between central features of the phenomenological method and the method of proofs and refutations: overcoming the separation of counterexample and the standard of knowledge, and the complementary changes occurring through criticism to the standards and objects of knowledge. According to Lakatos's arguments, the method of proofs and refutations inherently leads to skeptical opportunities within mathematical concept formation and theorem proving. From the perspective of uncovering the Hegelian content of Lakatos's work, this skeptical content is no coincidence, as Hegel's phenomenological method, on still closer inspection, is largely a transmogrification of ancient skeptical tropes.[5] Because that story is not well known, and it is so central to understanding the method of proofs and refutations, it is presented here. This latent skeptical background to *Proofs* will also provide the means to realize the promises of *Proofs* closer to the mathematical present.

Skeptical Foundations

A good place to begin is with Hegel's distinction between ancient Pyrrhonian methodological skepticism, as summarized in Sextus Empiricus's *Outlines of Pyrrhonism,* composed around 200 C.E., and the modern subjective skepticism associated with Descartes, Hume, or even Kant.[6] Cartesian skepticism assumes the position of a possibly deceived consciousness or self, and asks whether and how certainty in knowledge can be achieved through this quasi-psychological or cognitive agency. In Hume's account of causality and Kant's depiction of intuitions and understanding, the Cartesian idea again appears that skepticism is a threat to some type of interiorized subjectivity or cognition that may or may not be able to fend it off. A privileged subjectivity for Descartes or internal representations including synthetic a priori propositions for Kant are redoubts intended to protect the validity of perceptions and empirical knowledge generally from skeptical assault.

For Hegel, the modern expressions of skepticism, oriented around perceptual or subjective-internal criteria, are entirely different in approach from ancient skepticism's attacks on certainty of expressed knowledge claims of any kind. Pyrrhonian skepticism is not a doctrine about cognition and phenomenal appearances, as are several modern skepticisms or the responses to them, but a set of methods for showing that claims for certain knowledge, or for the essential nature or reality of any kind of object, will fail when criticized through a variety of skeptical heuristics. The limited negative goal of Pyrrhonism is to avoid affirming such certain beliefs. Cartesian-style skepticism and the Cartesian self have no more to do with ancient skepticism than, say, modern theories of meaning or reference that might also be used to justify foundational knowledge. No special status is assumed for individual subjective consciousness in Pyrrhonism, nor is Pyrrhonism directed only at appearances in the sense of modern perceptions or mental representations. Pyrrhonian skepticism is also not a denial that we know but a body of techniques supporting a practice for reasoning back from claims that knowledge is, or can be, absolute or certain. The Pyrrhonian skeptic does not want to say we are wrong about an underlying reality, or cannot know it at all, say, the cause of a medical ailment: that would amount to an overstatement of negative dogmatism that one absolutely cannot have some knowledge.

A classic historical object of skeptical redress is the claimed certainty associated with Stoicism's "clear and distinct impressions." The skeptic might point out the possibilities, for example, that almost identical objects, like twin siblings, may lead to indistinguishable impressions and hence mistaken judgments about which brother is actually, say, Socrates versus a supposed twin. Or the impressions of a drunkard may have all the features of the Stoic's "cataleptic" impressions for the drunkard, though they are clearly misleading in terms of an underlying truth. In this case, of arguing against Stoic certainty, the skeptic tries to take advantage of both the subject and object "sides" of the Stoic's epistemic claim: either the same correct impression is created by two underlying realities, or an incorrect impression is created by what would be taken to be an underlying reality. The Stoic has replies, of course, such as that truly distinct entities have some distinguishing characteristics, and something like a normal or healthy state of mind has to be assumed. If so, then the skeptic observes that a new criterion has come on the scene, the Stoic has taken a different step, and the debate continues with the new claimed standards for this Stoic theory of error. The skeptical gambits often sound trivial, like the repeated trope of looking for conditions for different appearances to different people or even to animals, yet they also turn out to be remarkably effective when skillfully deployed. In any case, the heart of Pyrrhonian skepticism is not its content—such as anti-Stoic arguments, which confound Stoic perceptual theory with the Pyrrhonist's methodological response to the Stoics—but just its method of Socratic-dialectical debate. There is no Pyrrhonist theory of perception in response, just a deflation of certitude. Pyrrhonist techniques are exercises in constraint, expressed as limitations on the certainty of knowledge.

What Hegel saw was that ancient skepticism was a methodological doctrine, not an epistemological one, and that the modern reinterpretation of Pyrrhonist technique completely vitiated its purpose. Like the Stoics, the major expressions of modern skepticism for Hegel ironically embody varieties of dogmatism as well. Either specialized assumptions of the self, perception, or cognition dogmatically explain the sources of skepticism, or that skepticism is not after all a danger to certain types of knowledge. Even Hume, who did not provide a foundational substitute for the skeptical problems he identified in causality and perception, implied that ordinary empiricism should provide beliefs of the world secure from skeptical doubt. But any quest for certainty is itself the problem not to be solved by Pyrrhonism but dissolved. Modern philosophy, for Hegel, misunderstands a practice intended to avoid just the dogmatisms it reintroduces.[7]

"Skepticism," wrote Sextus of this practice,

> is an ability to place in antithesis, in any manner whatever, appearances and judgments. We call it an "ability" not in any subtle sense, but simply in respect of its "being able." . . . In the definition of the Skeptic system

> there is also implicitly included that of the Pyrrhonian philosopher: he is the man who participates in this "ability."[8]

This ability, or its characteristic methodology, includes key methods such as invoking the important problem of the criterion: that a criterion of absolute truth either is assumed dogmatically, or refers to some other criterion and hence a potential infinite regress.

> In order to decide the dispute which has arisen about the criterion, we must possess an accepted criterion by which we shall be able to judge the dispute, and in order to possess an accepted criterion, the dispute about the criterion must first be decided. And when the argument thus reduces itself to a form of circular reasoning the discovery of the criterion becomes impracticable, since we do not allow them [the dogmatic philosophers] to adopt a criterion by assumption, while if they offer to judge the criterion by a criterion we force them to a regress *ad infinitum.*[9]

In the case of the Stoic, the problem of the criterion means that either the incorrigibility of perceptions as a criterion of truth is assumed, or the criterion is shifted to a theory of error and the "healthy mind." To enable the problem of the criterion to be invoked, the skeptic may lay the groundwork by showing that, say, the Stoic's cataleptic impressions are consistent with mutually inconsistent possible underlying realities: the veridical reality accepted by the Stoic, or the skeptic's Socrates-twin, or something different from the drunkard's distorted image. The mutually inconsistent claims constitute skeptical *isostheneia,* or equally strong arguments for a dogmatic claim and an alternative inconsistent with it, and starting from some proposal about some certainly true knowledge or its essential character: this is the "antithesis," as Sextus called it, of appearances and judgments. Isostheneia is close to the modern logical notion of undecidability: that for a sentence P in the language of theory T, P is undecidable with respect to T if neither P nor not-P is provable from T. In that case, both T + not-P and $T + P$ are consistent if T is consistent.

If proof-in-T is the basis for truth claims about the world, then it cannot be said that P is either true or false; judgment is suspended on P, and T cannot be a complete criterion of truth. In developing isostheneia and the problem of the criterion, the skeptic does not thereby positively reject the proposed standard as a standard of knowledge but carefully curtails its claim to absolute truth. As Sextus remarked, "We do not employ the arguments against the criterion by way of abolishing it but with the object of showing that the existence of a criterion is not altogether to be trusted, equal grounds being presented for the opposite view."[10] Isostheneia for the skeptic, similar to undecidability for the logician, does not establish that there is absolutely no criterion of truth but only that "nothing is absolutely a criterion of truth."[11] There are different tropes or gambits for skeptical process, organized into various lists, but a typical Pyrrhonist progression is: dogmatic claim → isostheneia → problem of the criterion → rejection of instance of claimed certainty. This simplifies a historically heterogeneous practice, yet it suffices for appreciating the logic of Pyrrhonist method in Hegel's historiography.

A skeptical goal, then, is to "suspend" beliefs about appearance, not taking them as either absolutely true or false. This is called the *epoche,* and there may follow with it a psychological state of well-being known as *ataraxia,* characterized by an absence of the distress associated with fundamentalist knowledge claims. For some skeptics, the goal of skeptical practice is ataraxia and the epoche is the means, while for others, the epoche might be a sufficient end. In either case, one is not left with a positive epistemological doctrine as a "solution" to the problem of certain knowledge; rather, a series of negative counterarguments are deployed only in response to the excesses of one dogmatism or another, whether it arises through medicine, psychology, physics, music, or any other branch of learning. Much of Sextus's texts consist of the same heuristics applied indifferently to different epistemic domains.

When the Pyrrhonist is questioned about the status of her own skeptical pronouncements, she can reply that her conclusions are framed only within the context of a dialectical debate between herself and an antagonist who claims certainty through some criterion of truth. The skeptic does not claim to establish isostheneia in opposition to, say, a Stoic criterion of truth but only to be able to develop isostheneia beginning immanently from Stoic assumptions, and thereby show that the Stoics' standard should lead her, on her own assumptions, to abandon her account of the real.[12] Skeptical arguments are propounded in a series of stages, and the truth of the dogmatic starting point is only conditionally assumed to carry to completion the construction of isostheneia, much like using mathematical assumptions. Sextus said, for example, that in adopting Stoic vocabulary and logic, the skeptic "Arcesilaus did not, in principle, establish any criterion; but those who think he did establish one ascribed it to an attack."[13] Carneades similarly is famous for arguing in a speech in Rome for justice one day and against justice the next day. His successor at the Academy in Athens, Clitomachus, said that he had never been able to decide what Carneades believed himself.[14] The "concessions" of dogmatism taken up by the skeptic can be so complete that they are indistinguishable from the opponent's: the skeptic is able to expertly reproduce, for purposes of isostheneia, the dogmatic criterion of truth—meaning its tenets, central concepts, and logical tactics. This stage setting for immanent criticism is just as crucial for the Pyrrhonist as the logic of isostheneia and the problem of the criterion.

When the skeptic speaks of appearances, or *phainomenai,* examples of Stoic perceptual impressions and the skeptic's arguments from Stoic assumptions may suggest that appearances and hence skeptical arguments are framed within a perceptual or cognitive theory, along the lines of modern skepticism. But the skeptical distinction between appearance and truth is a formal one about the status of the criterion of knowledge, and not a substantive distinction between ways of knowing. All kinds of knowledge appear for the skeptic, including the conflicting value judgments seen among different cultures, as suggested by Aenesidemus's tenth trope in the list compiled under his name; clearly, these comparative judgments are not appearances of sense. The verb *appear* is here a linguistic means to detach presuppositions or commitments about the truth or falsity of the concepts being used, while still being able to say what is expected in carrying the debate forward.[15] What appears are largely implicit or explicit methodological rules, or reasoning following such rules, and the modern tendency to interpret appearance perceptually is here just an anachronism.[16] Today, a knowledge claim is often recast as a "theory," and the Pyrrhonist performs a similar abstraction.

The inclusion of the skeptic's own statements within appearance leads to a great dilemma for the Pyrrhonists and their followers through the centuries. For the true Pyrrhonist, *her* criterion of knowledge is also supposed to be just appearance. Taking knowledge as appearance means to suspend judgment on the truth or falsity of a particular claim. That process on a case-by-case basis is unproblematic. But what does it mean to make such appearance and its affiliated suspension of judgment a general condition of one's beliefs? Whether that standard is possible holistically or for an entire life is not obvious.[17] For example, as Richard Popkin points out, after the rediscovery of Sextus's writings in the early Renaissance, both Catholics and Protestants made considerable use of them in religious arguments, and "The intellectual core of this battle of the Reformation lay in the search for justification of infallible truth in religion by some sort of self-validating or self-evident criterions," such as defining Scripture as the standard of truth and its own rule.[18] Such a gambit would be the way to stop the infinite regress of the problem of the criterion that both Catholics and Protestants exploited well. In using skeptical arguments in defense of their own dogmatisms, Protestants and Catholics saw also that the only evasion of the problem was some kind of self-validation. That "end state" of skepticism, or how a skeptic lives his skepticism throughout life, rather than just deploying it in rhetorical battle, was never adequately dealt with in the ancient texts, and reappears in the application of skeptical tropes, whether in the service of a dogmatism or not. Again, as Popkin says of the opposed parties during the Reformation, "The problem of stating the Pyrrhonian view without self-contradiction is one of the persistent problems recognized by the skeptics, and one of the continual answers offered by the opponents."[19] A solution is not obvious. Michel de Montaigne, himself adroit with Pyrrhonian arguments, remarked in his skeptical masterpiece *The Defense of Raimond Sebond* that Pyrrhonists need a "negative language" in which to express their criticisms without overstating them and solve the circular problem of having to judge a judging instrument by what it judges. How does one take advantage of skeptical criticism without succumbing to an incipient destructive nihilism?

Hume argued for the impossibility of complete Pyrrhonism and that skeptics suffered a kind of existential bad faith. Can one suspend judgment and live by appearances *always?* "We shall then see," quipped Hume in his *Dialogues concerning Natural Religion,* "whether you go out at the door or the window," as if the skeptic is indifferent to the truth of falling out a second-story window.[20] Another possibility considered by Hume was to look more closely at what happens to the Pyrrhonist's own skeptical arguments. As Hume observed on the power of skeptical criticism, "The understanding when it acts alone, and according to its most general principles, entirely subverts itself, and leaves not the lowest degree of evidence in any proposition, either in philosophy or common life."[21] At the same time, such subversion itself depends on rational principles to exist as an argument at all. It would seem that because skeptical arguments are constructed immanently, skeptical reasoning has to be at least as strong as that to which it is opposed.[22] Hume does not see this deferral of responsibility as providing a potential positive basis for rationality as traditionally envisioned. If the truth of the dogmatic starting point is taken to be thrown into doubt by skeptical argumentation, then such doubt extends for Hume perforce to the skeptic and completes the annihilation of reason from within:

> If the skeptical reasonings be strong, say they, "'tis proof, that reason may have some force and authority: if weak, they can never be sufficient to invalidate all the conclusions of our understanding." This argument is not just; because the sceptical reasonings, were it possible for them to

> exist, and were they not destroyed by their subtlety, would be successively both strong and weak, according to the successive dispositions of the mind.[23]

For Hume, the skeptical arguments *are* as strong as any rational one, and self-annulling. The exercise of reason to ground itself produces an inevitable cycle in which reason subverts itself into skeptical conclusions, which, if in turn attacked as being only as strong as the reason to which they apply, do not then weakly justify rationality, but instead "consume" it and skepticism as well. One does not metalogically escape from skepticism; reason is just avoided entirely. Or one ends in a kind of reason entirely unfamiliar to Hume.

Hume took ancient skepticism to its logical extreme and also denied the practical possibility of living out a skeptical life in which one really doubts or suspends judgment on all questions of truth. It was for this inability to live out the reality of Pyrrhonian arguments that Antoine Arnauld asserted that "Pyrrhonism is not a sect of people who are persuaded of what they say, but is a sect of liars."[24] And so too Hume, who likewise remarked that "the great subverter of *Pyrrhonism* or the excessive principles of scepticism, is action, and employment, and the occupation of common life."[25] But this does not then throw skepticism into doubt for Hume. It means that those who choose to inquire into the philosophical problems leading to skepticism are necessarily led to lead a double life: they hold, on the one hand, a theoretically correct rationalism-leading-to-skepticism, and on the other, a naturalistically conditioned dogmatism, equally correct in a pragmatism needed in order to carry on with the business of life. Roughly speaking, this is what Popkin calls Hume's "schizophrenic" skepticism.[26] "No philosophical dogmatist denies," wrote Hume in his *Dialogues Concerning Natural Religion,*

> that there are difficulties both with regard to the senses and to all science: and that these difficulties are in a regular, logical method, absolutely insolvable. No sceptic denies, that we lie under an absolute necessity, notwithstanding these difficulties, of thinking, and believing, and reasoning with regard to all kind of subjects, and even of frequently assenting with confidence and security. The only difference, then, between these sects, if they merit that name, is, that the sceptic, from habit, caprice, or inclination, insists most on the difficulties; the dogmatist, for like reasons, on the necessity.[27]

Pyrrhonism Become Historiography

We will return later to the examples of Carneades and the Humean schizophrenic. For the present, it is enough to recognize the complexity associated with the completion of skeptical practice, which leads back to Hegel and his appropriation of skepticism. For Hegel in the *Phenomenology* adopts skeptical methodology in his phenomenological method, but with a differing end. Indeed, the phenomenological method is nothing but a reorganization and generalization of skeptical tropes deployed as a theory of consciousness and turned into a historiographical method, or nothing less than Pyrrhonism become historiography. The appearance, then, of skeptical conclusions in *Proofs*'s mathematical history is the result of a skeptical deep structure serving both as the critical engine and means for reconstructing history. It is at just this point that Hegel and Lakatos join and reinvent a centuries-long philosophical tradition. How does this occur in Hegel?

First, Hegel framed his entire inquiry into forms of knowledge as stemming from the skeptical problem of the criterion:

> If this exposition is viewed as a way of *relating Science* to *phenomenal* knowledge, and as an investigation and *examination of the reality of cognition* [which it is not], it would seem that it cannot take place without some presupposition which can serve as its underlying *criterion.* For an examination consists in applying an accepted standard. . . . But here, where Science has just begun to come on the scene [that is, as appearance], neither Science nor anything else has yet justified itself . . . and without something of the sort it seems that no examination can take place. (*PhS* 52)

Hegel, just like a Pyrrhonist, did not invoke an external standard for his "Science," as he labeled it, or presuppose his Science to evaluate the numerous gestalts of knowledge taken up in his text. Yet how could Hegel have positively claimed to have developed his ramified knowledge series without violating the terms set by the problem of the criterion? How does one criticize a shape of knowledge and improve on it without assuming some other standard in turn, thus raising the problem of the criterion itself? Hegel's problem, expressed using the key terms and concepts of Pyrrhonism, is the perennial one of rendering a thoroughgoing Pyrrhonism coherent.

The first part of the answer lay in Hegel's characterization of consciousness as relational, which restates for the context of human consciousness generally what Sextus described as Aenesidemus's eighth trope: the so-called trope from relation. According to this trope, wrote Sextus, "we conclude that since everything is relative, we will suspend judgment as to what things are absolutely, and with regard to nature."[28] "In relation" means grounding knowledge alternatively from the perspective of the knowing subject and the object known, and both in relation to other things. The trope from relation, in contrast to the method of isostheneia, can be seen as a general scheme for organizing conflicting appearances involving relational predicates, such as "*x* appears to *y* as *p*."[29] In the case of the Stoic appearance of a Socrates-twin, the relational aspects of perception are exploited by the skeptic to create isostheneia and this intermediate role is typical in Pyrrhonist practice. Thus, the heuristic sequence is: dogmatic claim → relational analysis → isostheneia/contradiction → problem of the criterion → rejection of instance of claimed certainty. At the same time, the trope from relation is potentially stronger than isostheneia. Isostheneia only implies that *either* one or the other of opposing equipollent conclusions *might* be true. But identifying an intrinsic relation within a knowledge claim may imply that *neither* "term" of a relational condition can be a standard of truth on its own just because neither relata can meaningfully be separated out from the holistic relation. Hegel's solution to addressing the problem of the criterion vis-à-vis his proposed history of consciousness is to characterize consciousness as relational: "Consciousness simultaneously *distinguishes* itself from something, and at the same time *relates* itself to it, or, as it is said, this something exists *for* consciousness; and the determinate aspect of this *relating,* or the *being* of something for a consciousness, is *knowing*" (*PhS* 52).

What does relationality do for us? The intentional relationship between a knowing subject and the object known implies, Hegel said, that "consciousness provides its own criterion from within itself, so that the investigation becomes a comparison of consciousness with itself; for the distinction made above [that is, its relational character] falls within it" (*PhS* 53). Relationality, for Hegel, meant that the problem of criterion was at least temporarily evaded since a criterion of truth was implicit in knowing per se. This "comparison," as we already saw above in contrasting Hegel's method with that of proofs and refutations, is the extent to which the goals of a shape of knowledge are or are not met, either viewed from the perspective of the assumed criterion of truth, or taking the "object" as the criterion, just as the trope from relation may be deployed from its two sides. The two sides of knowledge claims in Hegel, also taken up by Lakatos as a proof and the objects to which it applies, are the relata of knowledge in general; but then, the criterion of knowing is implicit and need not be imported from without. For Hegel, consciousness generally allows for immanent critique whatever its modes, just as the skeptic immanently appropriates the concepts, vocabulary, and paradigm examples of the opponent, whoever he or she may be. "The essential point to bear in mind throughout the whole investigation," said Hegel,

> is that these two moments, "Concept" and "object," "being-for-another" and "being-in-itself," both fall *within* that knowledge which we are investigating. Consequently, we do not need to import criteria, or to make use of our own bright ideas and thoughts during the course of the inquiry. . . . [N]ot only is a contribution by us superfluous, since Concept and object, the criterion and what is to be tested, are present in consciousness itself, but we are also spared the trouble of comparing the two and really *testing* them, so that, since what consciousness examines is its own self, all that is left for us to do is simply to look on. (*PhS* 53–54)

This "looking on" is exactly skeptical appearance, phainomena, and represents much of the philosophical meaning of phenomenology in Hegel's title. Appearance was the next skeptical tool Hegel used, following naturally from his characterization of consciousness as relational. What mattered for Hegel was that we can reenact, from history, the same conditions the skeptic enacts in dialectical debate; this impersonation and even ventriloquization depends on criteria being implicit in consciousness, which in turn was a consequence of Hegel's nonmentalist intentionality, his relational consciousness. Ancient skepticism allowed for any topic-neutral appearance, including those of skeptical discourse, so Hegel's extension to *historical* appearances of knowledge was entirely legitimate and a remarkable innovation in the skeptical tradition. For example: "But Science, just because it comes on the scene, is itself an appearance; in coming on the scene it is not yet Science in its developed and unfolded truth" (*PhS* 48)—that is, the *Phenomenology* itself is to be taken as the appearance of Hegel's *Logic,* but the latter is not taken as a presupposition for the former. Or, "Since our object is phenomenal [skeptically appearing] knowledge, its determinations [specific shapes] will at first be taken directly as they present themselves; and they do present themselves very much as we have already apprehended them" (*PhS* 48)—namely, they appear to us as relational consciousness. Then, most important in deploying the equivalent of skeptical appearance: "Yet in this inquiry knowledge is *our* object, something that exists *for us*" (*PhS* 53), by which the skeptical impersonation of knowledge claims occurs as "our object," and we recognize it by letting it appear "for us," the phenomenological and narrative "we" appearing throughout the *Phenomenology.*[30] What appears for this "we" is just Hegel's rationally reconstructed history of consciousness. At the same time, this "we" is self-described by the *Phenomenology,* or at least that was Hegel's intent, and the "we" that has been interpreting the history of consciousness now also appears itself as a shape of knowledge. As Hegel simultaneously created the "we" of a generic historicizing sensibility, the *Phenomenology* constructs a "generic" appearance of modern consciousness that becomes one with the former as well. It is the relational and temporal structure of consciousness that allows this process to bootstrap itself through a history of the present without violating the problem of the criterion, or at least that was Hegel's aim and proposed solution. So far, Hegel's invocation of the problem of the criterion, his use of skeptical appearance as topic neutral and nonperceptual, and the approach of immanent criticism, with the latter following via a relational characterization of consciousness, are almost exactly Pyrrhonian.

Where, then, does Hegel start to differ? He differs little from the Pyrrhonian skeptic in the assessment of specific knowledge claims, but very much in the goals of those assessments, or what one does with skeptical conclusions—or "results" if "conclusion" is too strong.

The skeptic's typical product is an instance of isostheneia, or a pair of equally strong yet inconsistent extensions of a provisional dogmatic starting point. Hegel admired the ancient skeptics most of all for their methods for developing isostheneia as a discursive and learnable process for exposing the internal contradictions of knowledge claims. But isostheneia, as two mutually incompatible extensions of a dogmatic starting point, appears to have been too weak for Hegel, who wanted to identify more explicit contradictions for any shape of consciousness. That is, it was not sufficient for Hegel's purpose that he showed that a form of knowledge could be continued in either of two incompatible directions, but that *both* directions, and hence an inconsistency, are intrinsic to the knowledge itself. There is an ambiguous skeptical slogan, *ou mallon,* that refers to the twin sides of isostheneia as "this or that."[31] The weak reading of ou mallon is as merely *either* this or that—in the sense that either $T + P$ or $T +$ not-P could be true—but the stronger form, needed by Hegel and associated with the trope from relation, is *neither* this nor that—in the sense of an epistemological relation entailing two contradictory forms, one each for the subject and object sides of the relation. That, indeed, was the core of Hegel's account of error in the *Phenomenology,* that his shapes of knowledge contain internal contradictions and entail a stronger requirement than isostheneia by itself in order that the phenomenological "we" make progress. The shortcomings or errors in a shape of knowledge can be traced back to the intrinsic relations between subject and object that are characteristic of how that knowledge is defined. In principle, this stronger condition beyond mere isostheneia is made possible for Hegel by his assumption of the relational status of consciousness.

The next difference to consider is what Hegel did with the contradictions of a single gestalt of knowledge, and then the form of the *Phenomenology* as a whole. The unit of progress in the *Phenomenology* consists of two shapes of knowledge and a transition between them, and the entire book is the compilation of some score or so of such analyses. Here is where Hegel sharply changed direction from the ancient skeptics. Hegel took it as a shortcoming of skepticism that their refutations just peter out, that nothing is learned from skeptical assault notwithstanding the possible goal of ataraxia. This is exactly the issue noted earlier of "separating" falsifying counterexamples from the knowledge claim to which they apply; Hegel's lemma-incorporation, so to speak, was a response to skeptical perpetual searching.

For Hegel, the answer was to organize his patterns of knowledge so that contradictions uncovered in one are addressed and overcome by the next in his reconstructed sequence, while preserving what is worth keeping from the previous form of knowledge: this is the consistent improvement of Aufhebung, and the process continues for the entire *Phenomenology.* The sequencing of the individual patterns themselves also depends only on *our* historiographical reconstruction of the sequenced shapes: the sequence is not given automatically, and is not the product of a historical demiurge or other mysterious process. As Hegel stated, "This way of looking at the matter is something contributed by *us,* by means of which the succession of experience through which consciousness passes is raised into a scientific progression—but it is not known to the consciousness that we are observing" (*PhS* 55). Hence we, as historians and self-constituting modern selves, are the purposeful creators of this rational reconstruction, in which historical patterns of knowledge are organized into a pedagogically useful and critical history of consciousness. This method, however, just defers raising the problem of the criterion. Either judgment is again suspended on each new shape of knowledge or criticism of it continues. Where does the process stop?

Hegel's goal was not skeptical ataraxia but historical self-consciousness in what he considered to be the "completion of the series," which has embedded in it *all* the characteristic philosophical positions provided by Hegel in his book, as well as their internal contradictions and some means of addressing them. That is, in the words Hegel used to describe the form of knowledge he called Religion,

> The "*shape*" in which it appears for its consciousness will be perfectly identical with its essence, and it will behold itself as it is. In this genesis of religion, Spirit itself therefore assumes *specific* "shapes" which constitute the different moments of this movement. . . . [I]ts consciousness and its self-consciousness are on the same level. . . . This self-consciousness has, as consciousness, itself for object. (*PhS* 414–15)

The overall trajectory of the *Phenomenology* is from individual knowledge through forms of social organization and communal knowledge, such as the Greek polis, feudalism, and the Enlightenment, then back to the reflective knowledge a culture has of itself through art, religion, and philosophy; the holistic and totalizing features of the last shapes of knowledge then make plausible Hegel's idea that they overcome subject-object dualisms. Leaving aside whether that is truly achieved, or the possible shortcomings of a view of human history or behavior in which it might be claimed to be achieved, the point here is that what Richard Popkin refers to as the problem of a "self-validating" criterion, and the problems seen for hundreds of years in living out skeptical claims, Hegel attempted to solve by constructing an epistemological theory that becomes its own metatheory, to use contemporary parlance. That was Hegel's end for his Pyrrhonism become historiography.

There is one last skeptical technique to be identified here that will figure in the sequel. In this "way of looking"—each piecewise step to the next minimally improved conception of truth—just what happens at each change from one pattern of knowledge to the next? Hegel said that "from the present viewpoint [that is, his and the reader's], however, the new object [shape of knowledge] shows itself to have come about through a *reversal of consciousness itself*" (*PhS* 55), meaning that each shape of consciousness becomes in some way the object of its own knowledge and to that extent expressive, however primitively, of self-consciousness. In other words, "our knowledge of the first object, or the being-*for*-consciousness of the first in-itself, itself becomes the second object" (ibid.): this is the "reversal." Thus, Hegel's heuristic sequence is: dogmatic claim → relational analysis → isostheneia/contradiction → problem of the criterion → rejection of instance of claim → reversal and invocation of new standard.

Even this is still within the orbit of skepticism, as is prudent, or else Hegel would be open to requiring an additional justification, however weak, for his choice of contiguous shapes. Hegel's reversal was just the skeptical *peritrope* or "turning of the tables" of a criterion of knowledge back on its proponent, such as Plato's description in *Theaetetus* of Socrates's turning the "man is the measure" doctrine back onto its proponent Protagoras. Characteristic of the peritrope is not just its conceptual turnabout but the way in which it points back to previous steps, stages, or argumentation, and includes the performance of these stages within the scope of the conceptual reversal. In ancient times, the concrete circumstances in which the peritrope was applied typically would be an oral debate between two opposed speakers: not a unified abstract argument but a sequentially articulated series of moves and countermoves.[32] The invocation of the peritrope brings about the conceptual self-application of an argument as it points to the specific delivery of the argument in speech, so that the peritrope functions as the oral equivalent of critically deployed or perhaps ironic quotation. It is the realization of self-application that gives an instance of the peritrope its point, not just the reversal *in abstracto,* as the skeptic empirically identifies the appearance of a critical standard in the speech of an opponent; Hegel looked back at both empirical history and the *Phenomenology* itself. The peritrope is then uncoincidentally at the heart of the problem, perceived so accurately by Arnauld and Hume, of how the skeptic lives out a skeptical life, since the peritrope provides the internal starting point for a critique of skeptical practice; but as mentioned above, at least Hume felt that such a critique was not entirely possible since it led to a disintegration of reason instead of reason's rescue.

The peritrope also has positive uses, meaning that the "turning of the tables" effected by the reversing gambit may show that in the execution of the argument, the speaker of the argument does meet the criterion of truth used as his or her own. Pyrrhonists, may, for example, have believed themselves immune to such a reversing attack in declaring that their epistemic criterion in life was appearance. For Hegel, because of the inclusiveness of his series of patterns of knowledge in the entire *Phenomenology,* its entire pattern, he implied, can safely be "refuted" because the refutation should already be embedded somewhere in the whole series. Whether such logical inclusiveness can be demonstrated or not is obscure given the overwhelming complexity and inclusiveness of Hegel's book. But the motivation is clearly that Hegelian historical Science may conceptually "reverse" itself, and that *that* is part of its pattern of knowledge. The presentation of the entire series of shapes of knowledge still is appearance—namely, the historical appearance of Hegel's philosophy itself—and therefore, makes no additional claim to a more fundamental metaphysic or standard, save perhaps that this appearance is of Hegel's present moment in history. Instead of the epoche and suspending judgment on truth and falsity, Hegel reinvented rationality as a cognitive and historical process for which truth and falsity become relational terms within an elaborate scheme of conceptual change. That goal is what appears in the *Phenomenology.*

The story deserves continuation. But because *Proofs* does not share Hegel's goal of a cumulative, self-refuting, though internally consistent system, this account of the skeptical techniques embedded in the phenomenological method can come to an end. The description so far of Hegel's Pyrrhonism become historiography is sufficient to complete the look at Lakatos's mathematical skepticism below and to apply it in chapter 4.

The End of Mathematical Skepticism

We saw that the method of proofs and refutations is strikingly similar to Hegel's phenomenological method, through Lakatos's idea that lemma-incorporation builds a counterexample into the structure of a theorem or its lemmas, and that similarly enriching transformations in mathematical content could occur in changing conceptions of the mathematical objects a theorem is about: such is the intrinsic unity of the method of proofs and refutations, and also Hegel's notion of epistemic progress by which he contrasted his historical reconstruction with the disorganized eclecticism of classical skepticism. The process on which criticism of a theorem–conjecture–thought experiment is predicated, Lakatos argues, was the possibility of concept stretching as the mechanism for refutational success, which for him, in turn, makes possible unmitigated skepticism about any informal mathematical proof. Modern logical theory—primarily theories of logical constants and quantifiers, and proof and model theory—from this perspective, provides a form of a mitigated skepticism that has stabilized mathematical criticism for the sake of theoretical progress. Recognizing Hegel's phenomenological method as a historicized conception of Pyrrhonian skepticism helps to identify more precisely how the classical skeptical apparatus is realized in *Proofs* or contemporary mathematical methodology itself.[33] Because Lakatos is explicit about the skeptical consequences of lemma-incorporation, it is enough to make brief observations here on the principal components of skeptical technique, and then move on in the next chapter to a detailed look at its mathematical consequences.[34] These observations will complete the skeptical framework created in *Proofs* for the production of mathematical knowledge.

First, skeptical appearance and immanent reason. For the purpose of creating mathematical knowledge, does a mathematician, or anyone else, need to believe in the substantive truth of mathematical theorems, or is it sufficient to consider theorems as inevitably of the form $P \rightarrow Q$, with mathematical conclusions having the same provisional status of conclusions in a skeptic's immersion in the conditional assumptions of an opposing dogmatist? Today, it is a commonplace that it is largely immaterial whether the axiomatic starting point in the practice of mathematics is "really true," as used to be thought of both arithmetic and Euclidean geometry. How could one *do* both Euclidean and non-Euclidean geometry, or commutative and noncommutative algebra, without being able to suspend judgment about the absolute truth of one's staring point? Not doing so would be irrational. It is not as if these contradictory options occur within a completely stable setting—say, one made available through some kind of set theory to represent all of mathematics, since the process occurs there as well. The beginning of the twentieth century saw many attempts at creating *a* foundation for mathematics, meaning in part to represent the corpus of mathematical knowledge as being generated from a finite or, more generally, recursive set of axioms by the application of a finite set of precisely specified rules of inference. Bertrand Russell and Alfred North Whitehead's *Principia Mathematica* is perhaps the most famous such system, but today a good student of mathematical logic can enumerate a half-dozen candidates for mathematical foundations, all having their own exotic but mutually inconsistent consequences for various mathematical conjectures. It is possible, for example, for texts in mathematical logic to pass through successive chapters, each one on a distinctive set-theoretic framework for the whole of mathematics, without the author having as much as to pause to justify such a mind-boggling theoretical stance.

Today, the advanced student of foundations must be equally at home in Gödel's "constructible universe" of sets, or in the so-called "playful universe" based on a formalized treatment of a game-theoretic interpretation of quantifiers, or with a cornucopia of axioms regarding the range of allowable infinities. In practice, one suspends belief on the ultimate status of these axiom systems, but in order to investigate them, assumes them as a criterion of truth and works within them immanently. One may object with the value judgment that such a skeptical attitude is somehow unsatisfactory, that there should be just a single set theory, just as it was once thought that there existed only one geometry, Euclidean geometry. Yet for mathematical practice, such a value judgment today has as little force as would an injunction to limit oneself to a single type of geometry or algebra.[35] Standards of progress may be invoked from time to time, such as judgments as to whether a field of inquiry is effectively exhausted, or no longer delivering interesting or useful results. But that does not mean that one part of mathematics is accepted as true and another false in any fundamental sense. As a practice, theorem proving does appear much as skeptical immersion, like provisionally assuming a Stoic theory of truth, but then withdrawing complete assent in moving on to another topic of study.

Next, *Proofs* provides an inventory of ways in which the skeptical problem of the criterion and its generic antifoundationalism can be raised about informal proofs. The method of proofs and refutations in general, and concept stretching in particular, are means for exposing the problem of the criterion in specific mathematical contexts. For contemporary mathematics, theorem-proving practice appears not to use or need a foundation of mathematical truth, either of what numbers, sets, or functions really are, or of theorems that are unconditionally true. Much is known in mathematics about relative consistency, meaning that one can prove that if theory T is consistent, then so is theory T', or that T' can be represented within T itself, say by "reducing" numbers to sets or vice versa. This ubiquitous conditionalization of mathematical knowledge is one of the most salient lessons of modern logic. To attempt to evade it means invoking, from a skeptical perspective, some dogmatic standard of the human mind, the mathematical needs of physics, or some other such criterion of truth to make the conditioned unconditioned. No such attempt has yet been successful, nor appears to be required for mathematical practice. Not even conditional mathematical truths of the form $P \rightarrow Q$ are absolute in *Proofs* since they, too, are part of fallible mathematical practice. Lakatos's clearly expressed position about informal mathematics generally, regardless of the varieties of set theories and alternative "foundations," is that refutations are always possible, and that neither certainty nor foundations are ever found. The conditions supporting that conclusion are just those marked off by developing the problem of the criterion.

Then there is isostheneia as well as the trope from relation, which can be taken together here. One affinity between Hegel's phenomenological method and that of proofs and refutations lay in how counterexamples become embedded in proofs and proofs drive changes in domains of mathematical objects. In Hegel, that process was at bottom an expression of the relational character of consciousness. The power of relational definitions for criticism is translated through lemma-incorporation into the relational character of informal mathematical proofs: that proofs are about mathematical domains just as consciousness is always consciousness of its objects of knowledge. The power of lemma-incorporation and concept stretching is an instance of the more general power of criticism exploiting the problems of securing a foundation for knowledge using relational definitions. Lakatos argues that lemma-incorporation is the engine of mathematical progress, but that basis now turns out to be another aspect of skeptical criticism. The method of proofs and refutations contains, or simply is, a sophisticated form of the trope from relation, which in conjunction with the absence of any fundamental standard of mathematical truth, expressed through various forms of the mathematical problem of the criterion, makes possible indefinitely continued mathematical refutation. And just as for Hegel isostheneia is weaker than outright contradictions in a form of knowledge, so too, Lakatos needs refutation to generate progress in informal mathematical theories. This subtle though strong shift makes possible conditional progress, through a superseding shape of consciousness for Hegel or an improved theorem for Lakatos.

The underlying structure of Lakatos's heuristic toolkit, therefore, is composed of skeptical ideas, especially as refracted through Hegel's phenomenological historiography: the ability to completely take up an epistemic stance without committing to its fundamental truth; the trope from relation; the possibilities for subject-object concept stretching in informal mathematical discourse; and the potential for always raising the problem of the criterion. Hegel combined skeptical criticism within a narrative form dominated primarily by choices of scale and time. In mathematics, Lakatos uncovers its skeptical core as the problem of balancing criticism with the dangers of unmitigated skepticism.

Where *Proofs* differs considerably from ancient skepticism and Hegel is in the goal or end state of these implicit skeptical practices. Hegel's aim, as we saw, was the reflexive, self-accounting shape of modern historical consciousness. The most that should be taken from that for *Proofs* is that Lakatos has provided, as I have claimed all along, a classic account of the constructive historical status of modern mathematical method. It now turns out that the logic for that self-generative activity, even its Bildung, is mostly contained in the tropes of ancient Pyrrhonism. *Proofs* opts for a mitigated skepticism, of provisionally accepting some conception of mathematical truth—of the roles for logical constants, of logical versus extralogical content, on context-independent methods of proof—that ultimately is subject to skeptical attack, but that would also destroy or radically reinvent the subject as we know it. As Pascal said in defense of mitigated compromise, "We have an impotence to prove, which cannot be conquered by any dogmatism; we have an idea of truth which cannot be conquered by any Pyrrhonian skepticism."[36] Such mitigated skepticism is continued by Lakatos, and is more than a credo; it can be seen in the method of lemma-incorporation and features of contemporary mathematical practice generally.

Unlike Hegel's philosophical end, Lakatos's mitigated skepticism is largely a pragmatic choice driven by the external goal of maintaining a coherent body of mathematical knowledge. That naturally reflects value judgments about the desirability of mathematical progress, just as the Pyrrhonist makes value judgments about the desirability of the end state of ataraxia, or even about just jettisoning a dogmatism via the epoche. In the historiographical framework of *Proofs,* skeptical chaos is tempered through the constrained emergence of the present from the past. People make their own mathematical history, but not entirely as they please. Mathematical skepticism is guided by the practical goal of improving mathematical knowledge: extending the scope of theorems, improving their rigor, formulating novel conjectures, and creating informative new proofs for old theorems. That purpose is wholly analogous to the aims of constructive skepticism in science, whether its seventeenth- or twentieth-century versions. Mathematical skepticism is not just a bothersome theoretical inevitability; it is a useful engine of conceptual progress that just has to be restrained lest mathematics descend into chaos. Lakatos's mitigated skepticism is just Humean schizophrenic skepticism, in which practically one makes provisional dogmatic assumptions to serve life's goals or mathematical ones, while theoretically maintaining, as Pascal also said, "*Le Pyrrhonism est le vrai.*" The air of paradox is dispelled by the historical view of *Proofs:* that radical, unmitigated skepticism always can occur, but only does occur, or is allowed to, when needed to reorganize our knowledge in major ways, as happened in mathematics at the turn of the twentieth century. Hume's schizophrenia, properly administered and modulated, can be a productive one.

Such are the roles of skepticism in *Proofs:* skeptical techniques and attitudes for modern mathematical practice, slanted, as for Hegel, toward—fallible!—refutation rather than the mere undecidability of isostheneia; a shared antifoundationalism via the problem of the criterion; and the use of pragmatic goals different from the epoche or ataraxia. That last transformation of skepticism, of setting a different goal, is similar also to what occurs in Hegel, where he adopted the goal of self-validating knowledge, or a reflexive metatheory, and which is absent from *Proofs.* In both the *Phenomenology* and *Proofs,* the shared pedagogical and historical presentation turns Pyrrhonism into historiography and historical criticism. Now that the secret of concept stretching's success is clear, the next step is to pursue skepticism further into the details of *Proofs* and its role in what Lakatos takes to be the locus of real refutational potential in contemporary mathematics, being the boundary between informal and formal mathematics and the translation process between the two.

This page intentionally left blank

4 | Between Formal and Informal

In chapter 2 of *Proofs,* Lakatos discusses the problem of translating an informal proof into a mathematical theory with "perfectly known terms" and "infallible inferences" (106–26).[1] The example Lakatos uses is the translation of Poincaré's 1899 proof of Euler's theorem into the terms and arguments of vector algebra: relations between the vertices, edges, and faces of polyhedra become relations between various vector spaces represented through the incidence matrices of the polyhedra, meaning matrices of 0's and 1's representing which vertices belong to which edges, then which edges to faces, and finally which faces to the polyhedron. Euler's theorem follows from these defined relations and standard results relating the dimensions of associated subspaces and the ranks, or vector-space dimensions, of the matrices representing them. Dating from the close of the nineteenth century, this proof of Euler's theorem is a fitting conclusion for Lakatos's mathematical bildungsroman.

The central figure here is Epsilon, the speaker for what Lakatos calls the "Euclidean program." For Epsilon, there are "Euclidean theories with perfectly known terms and infallible inferences—like arithmetic, geometry, logic, set theory . . . and he now makes the Euclidean program consist" of translating formal mathematical theories into such available, established, "already Euclidean" or dominant theories (*P&R* 123). Within his dominant theory, Epsilon claims infallibility. Once, say, an informal argument of probability theory or the theory of computation is translated into the calculi of probability axioms or Turing machines, Epsilon alleges certainty for his proof. Lakatos will neither question nor endorse this claim made from inside a formal or neoformal system, but instead sees the process of formalization as subject to falsification and challenges what certainty inside a formalism is certainty of: whether certainty of an informal mathematical conjecture or object of pure syntax. Epsilon virtually translates Euler's theorem into first-order predicate logic so that his claim represents the position that by twentieth-century standards, as Poincaré himself wrote, "absolute rigor is attained."[2]

But by the close of the chapter, Epsilon has retreated from a claim to grasp the essence of modern informal notions of point, edge, connectedness, boundary, and so on in his dominant algebraic theory. Epsilon admits that "point," "vertex," "edge," and so forth can only function as formal abbreviations within his dominant theory for him to assert absolute certainty and finality for his proof that "not even a monster [counterexample] can refute" (*P&R* 124). In moving back and forth between informal and formal proofs, Lakatos argues that Epsilon's claim to certainty entails severing all ties to informal reasoning, which exist through an equivocation in the Euclidean's definitions. Whatever certainty is achieved by the Euclidean depends on interpreting the "is" of definitions ("an edge *is* . . .") purely as an abbreviation within the formalism, not an explanation of content. But to attribute any meaning to Epsilon's proof, the "is," in at least some cases, must be taken in an essentialist or predicative sense, as translating some informal property into formal terms. The proof is certain, or as certain as can be, Lakatos contends, as long as all copulas are interpreted as identity, yet the proof only becomes meaningful when the copula "is" is interpreted as predication. The informal component may be jettisoned, although only at the cost of transforming the formal proof into a purely syntactic object that is either devoid of meaning or may misrepresent aspects of the original; metamathematical and logical concepts, too, are not excluded from this possibility. As put by Alpha,

> If you regard your translation as infallible, or, if you consciously scrap the old meaning, both these extremes will yield the same result: you may push out the original problem into the limbo of the history of thought—which in fact you do not want to do. . . . [Y]ou have to admit that [a Euclidean] definition must have a touch of modified essentialism: it must preserve some relevant aspects of the old meaning, it must transfer relevant elements of meaning from left to right. (*P&R* 122)

Then, as Gamma expresses the problem:

> I am a bit puzzled by your definition of polyhedra. In the first place, as you bother to define the notion of polyhedron at all, I conclude that you do not consider it to be perfectly well known. But then where do you take your definition from? You defined the obscure concept of polyhedron in terms of the "perfectly known" concepts of faces, edges and vertices. But your definition—namely that the polyhedron is a set of vertices, plus a set of edges, plus a set of faces, plus an incidence matrix,

> obviously fails to capture the intuitive notion of a polyhedron. It implies, for instance, that any polygon is a polyhedron, as is, say, a polygon with a free edge standing out of it. Now you must choose between two courses. You may say that [quoting Pólya] "the mathematician is not concerned with the current meaning of his technical terms. . . . The mathematical definition creates the mathematical meaning." In this case to define the notion of a polyhedron is to drop the old notion altogether and to replace it by a new concept. But then any resemblance between your "polyhedron" and any genuine polyhedron is entirely accidental, and you will not get any certain knowledge about genuine polyhedra by studying your mock-polyhedra. The other course is to stick to the idea that definition is clarification, that it makes the essential features explicit, that it is a translation or a meaning-preserving transformation of a term into a clearer language. In this case your definitions are conjectures, they may be true, they may be false. How can you have a certainly true translation of a vague term into a precise one? (*P&R* 108)

Lakatos's exposition of Poincaré's proof shows that the Euclidean does not argue within her formalism but between the formalism and informal arguments that interpret the formalism. The informal component can be abandoned completely, but then nobody will recognize the formalism as mathematics.[3] The Euclidean's formal program is essentially tied to informal mathematics by informal translations, and says Teacher, "I indeed think that this method of translation is the heart of the matter of certainty and finality of Epsilon's proof" (*P&R* 121).[4] The characteristic feature of the translated proof is that its rigor does not hinge on terms specific to linear algebra, such as vector, subspace, incidence, and so on. Anything—in particular, edges, surfaces, and polyhedra—can be appropriately substituted for these specific terms and there will be no refuting counterexamples, in contrast to the decades of conjectures and refutations preceding this proof. But we may not substitute what we please for nonspecific terms like the logical connectives "and" and "not," or quantifiers and natural numbers and sets: that kind of substitution is ruled out by fiat. In this interpretation, as discussed earlier in chapter 3, Epsilon's proof is an instance of Bolzano's relativized logical truth with which Lakatos closed out his pre-Poincaré account of Euler's theorem. The proof is irrefutable because it does not depend at all on the interpretation given to nonspecific terms, at least within the confines of the underlying elementary set theory that provides a well-defined universe for the range of predicate and individual symbols. The certainty achieved by Epsilon through Euclideanization is a result of restating all aspects of the theorem to be proved in perfectly known terms that are completely vacuous given the non-specific logical theory. You cannot destroy your meaning and have it too.

The burden of infallibility, therefore, is also carried solely "by the meaning of the non-specific, underlying terms—in this case arithmetic, set theory, logic" (*P&R* 124). Formal proofs "do not depend at all on the meaning of the specific terms," and "by pinning down their [the nonspecific terms] meaning we state what can be accepted as counterexamples and what cannot. Thus we regulate the spate of counterexamples" (ibid.). That is, Epsilon can even say inside the formalism—with as great a certainty as one has of anything—what could count as a monster or counterexample. Still, as Beta suggests, Epsilon's translations "seem to appear as mere abbreviations in the dominant theory and thus they are true 'by definition.' But they seem to be falsifiable if we regard them as referring to the non-Euclidean [that is, informal] realm" (*P&R* 123). So Lakatos chooses not to contest Epsilon within formalism, but redirects attention to Epsilon's subtle relationship to his own informal mathematical activity. Particular Euclidean proofs may be "falsified" if some of their consequences lead back to unacceptable informal results. For example, in Poincaré's proof, a basic concept is that of the boundary. The boundary of a polyhedron's face is basically the edges belonging to the face, and the boundary of an edge is the set of its vertices. For the proof to work, the empty set is defined as the boundary of a single vertex: the empty set is to a vertex as a vertex is to an edge, or an edge is to a face. As well, vertices could be defined as having *no* boundary, in which case the proof proves that $V - E + F = 1$ for the tetrahedron. Such a "proof" is falsified by the informal decision to maintain $V - E + F = 2$ for the tetrahedron, which of course should be done, but that also shows that a "modified essentialism" is maintained to select formal definitions.

The most extreme result from falsifications of this kind is that the dominant theory is modified or replaced—that is, one Euclidean theory can be replaced with a more desirable one. One set theory is replaced by another, or one version of probability theory, as the formal representative for informal notions of chance, is exchanged for another. But within such dominant theories, Epsilon can still claim certainty for his proofs because the nonspecific terms have been left untouched. Epsilon's absolute certainty, being close to contemporary standards of rigor, will become problematic only if his non-specific terms, such as the usual notions of proof from first-order logic or our understanding of sets, are subject to falsification, counterexamples, or monsters. As Lakatos notes, dominant theories have been scrapped in the history of mathematics: the Greeks substituted arithmetic with geometry, Descartes replaced geometry by algebra, Cauchy and Weierstrass later reintroduced arithmeticization, informal probability theory was embedded into measure theory, and "Russell in turn made logic the dominant theory of all mathematics" (*P&R* 125n1).[5] Russell's program, from this perspective, was the translation of all informal mathematics into Euclidean theories based on set membership ("ϵ") and equality ("="). As a translation between informal and formal mathematics, this means again, as Lakatos maintains, that "validity of an intuitive inference depends also on the translation of the inference from ordinary (or arithmetical, geometrical, etc.) language into the logical language: it depends on the translation we adopt" (*P&R* 123n2).

If formalization, however, is itself the potential locus of falsification, yet the target Euclidean theory just *is* what is meant by a formalized theory, how would a problem in translation show up? Could *that* dominant theory be shown as fallible in some meaningful way? In the introduction to *Proofs,* Lakatos says that like the conjectural versions of Euler's theorem, modern metamathematical theories are "a paradigm of informal, quasi-empirical mathematics just now [circa 1960] in rapid growth," and that "the student of recent history of metamathematics will recognize the patterns described here in his own field" (*P&R* 5). So formalism is not jettisoned but rather understood as an abstraction within Epsilon's informal mathematics, much like other mathematical objects. Only now, the exquisite feature is that the abstraction happens to be of mathematical knowledge itself. Euclidean proofs are themselves more formal objects like polyhedra, and as far as they represent mathematics, depend on informal, fallible translations. Just what examples, though, get at the heart of formalism as the dominant informal theory of contemporary mathematics?

Lakatos himself points out that even Gödel's famous incompleteness theorems, the epitome of modern mathematical logic, are themselves informal.[6] The suggestion is intriguing, as Gödel created the basic theoretical apparatus, now known as Gödel numbering, by which proofs are translated into statements about numbers. So if Lakatos's ideas have merit at a more detailed mathematical level, it is here we might begin. Our look will be both historical and, following Lakatos's lead, skeptical in interpreting the informal content of Gödel's work.

What Gödel's Proofs Prove

In 1931, Gödel proved his two incompleteness theorems that apply to axiomatic systems like Russell and Whitehead's *Principia Mathematica,* but his methods are sufficiently general to apply to many axiomatized systems, even weak ones. The first incompleteness theorem says, in effect, that a consistent axiomatic system strong enough to prove some weak theorems from elementary number theory—requiring only the operations of addition and multiplication, but neither operation separately—will be *incomplete:* there will always be mathematical sentences formulated in the syntax of the system under consideration that are neither provable nor refutable in the system, and these sentences are said to be undecidable with respect to the system.[7] The second theorem says that for the same type of systems, if they are consistent, then a formalized expression of the system's consistency is itself not provable in the system as well. The second incompleteness theorem, therefore, expresses the internal unprovability of consistency.

Since an undecidable proposition and its negation are each separately consistent with the base system, the old system can be extended to two mutually incompatible new ones by adding on the undecidable sentence, or alternatively its negation, as a new axiom. The classic example of this procedure is the generation of non-Euclidean geometries by adding the negation of the parallel postulate to the axioms of elementary geometry without the parallel postulate. But the new arithmetic systems so constructed also have new undecidable sentences, different from the originals, and the process of constructing new undecidable sentences and then new systems incorporating them or their negations goes on ad infinitum, like a branching tree that never ends. That is, Gödel not only proved these results for the axiomatic system he used; he provided a constructive process that applies to extensions of the base system created by adjoining the undecidable sentence or its negation as well. It can even be proved that there is a continuum of mutually inconsistent theories all containing, say, the axioms for Peano Arithmetic: in other words, as many inconsistent theories as there are real numbers. Gödel's first theorem showed that undecidability was endemic to much mathematical reasoning, and a major portion of foundational research since the 1930s involved exploring mathematically important conjectures that are undecidable with respect to various "foundational" systems—the ironic quotation marks now being mathematically necessary.

Now, a significant question is how far Gödel's findings are reflective of mathematics outside of formal systems. Gödel's original undecidable sentence is of no known mathematical interest apart from its undecidability in formal arithmetic and primary role in logical research, and for many years, it was an open problem to find a nonmetamathematical sentence similarly undecidable in Peano Arithmetic, the first such example being found in 1977.[8] Another line of research following from the model set by the first theorem turns the undecidability issue around. Why not investigate alternative mathematical systems that do decide some of the interesting undecidable propositions one way or the other?[9] Within any one of these systems, one can investigate those new theorems that *are* provable or refutable though undecidable with respect to relatively lean systems such as Zermelo-Fraenkel set theory without the axiom of choice. Cantor's continuum hypothesis, regarding the order of infinity of the real number line, is the most famous undecidable statement investigated in this way, but there are today several such propositions of mathematical, and not only metamathematical, interest. One strategy, then, in post-Gödelian foundational studies is reflected in these attempts to prove to mathematicians that metamathematics has a direct bearing on concrete mathematical problems, yet this progress has been made mostly across alternative foundational theories, primarily ones with specialized consequences for real analysis and properties of the real number line. This state of both logical and metamathematical theory is neither problematic nor catastrophic for logicians or mathematicians, and it is consistent with Lakatos's view that formalisms can be falsified, or at least preferred, because of informal choices made about their consequences. That was Gödel's own methodological approach, which influenced several generations of set theorists. But as indicated by Lakatos, such informal choices may be localized within Gödel's two theorems themselves.

Where can guidance be found? Lakatos's account in *Proofs* implies that a problem about formalization could appear as a type of skeptical dilemma regarding formal translations themselves, perhaps reflected in the heuristic structure of formalist concepts or the proofs in which they occur. Indeed, even the cursory statement here of Gödel's theorems and their aftermath is enough to suggest a skeptical account of the heuristic structure of his theorems.

Gödel's second theorem states that the consistency of a system under consideration, say *Principia Mathematica,* is itself not provable in that system, unless the system is inconsistent, in which case any sentence is provable. If the formalism is thought to be an adequate representation of mathematics as a whole, this means that an internal, mathematical justification is lacking against the possibility of proving "1 = 2," or absolutely *any* sentence formulated in the syntax of the system. Mathematical dogmatism, in a nonpejorative sense, here means that within a given formalized mathematics, *only* a specific set of sentences is provable and absolutely no more. From a skeptical stance, the natural interpretation of the second theorem is that it expresses the problem of the criterion for the absolute correctness of mathematical proof, in the system at hand, taken as the dogmatist's criterion of mathematical truth. The skeptical problem of the criterion asserts what Gödel showed can be established mathematically: either the consistency of mathematics must be assumed outright or invoked by some extramathematical standard, or we are forced into an infinite regress of foundational theories. To invoke a Kantian standard, for example, that the truths of mathematics are guaranteed by the transcendental structure of sensible intuitions of space and time, is, for the skeptic, to introduce a second standard that now guarantees the truth of the first, and the skeptical reply is to ask again whether this standard is self-validating or if it, too, depends on some further criterion of truth. Whether one thinks the problem of the criterion is valid in general is here irrelevant, since Gödel demonstrated with precision that for his proof systems, their internal consistency could be proven only by appealing to an alternative standard of truth whose strength, relative to the first, could be stated in mathematically exact terms. Gödel's second theorem proves that there is a rigorous interpretation of the problem of the criterion when the standard of mathematical proof is used as an absolute criterion of mathematical truth. And like the problem of the criterion, Gödel's proofs have complete generality: they may be applied to all kinds of mathematical systems with the same consequences. Stated positively, the consistency of mathematics is a conjecture that may be falsified by the discovery of a proof of inconsistency.

So far, this mostly amounts to a restatement of the standard interpretation of the second theorem, but the reformulation points to another aspect of Gödel's work describable through Pyrrhonist methodology. While the second theorem may be taken as expressing an instance of the problem of the criterion, the undecidability conclusion of the first incompleteness theorem is a version of skeptical isostheneia, meaning the construction of two alternative yet mutually inconsistent characterizations of the real or what is truly the case. For us, this means provisionally taking the proofs, say, of Peano Arithmetic as the truths of arithmetic, or the proofs of *Principia,* Zermelo-Fraenkel set theory, Zermelo-Fraenkel plus the axiom of choice or the axiom of determinacy, and so forth, as the truths of a universe of sets: "the real world," as set theorists sometimes say, though perhaps with some irony. As long as the base system is consistent, both the undecidable sentence and its negation are consistent with that system, and thereby provide two versions of "the real," which is inconsistent with the proof system considered as a standard of *truth;* that establishes Pyrrhonian isostheneia. As with the problem of the criterion, Gödel showed that the skeptical technique of constructing isostheneia is rigorously possible within mathematics when certain formal standards of proof are put forward as informal standards of mathematical truth. And the skeptical trope of isostheneia, like the problem of the criterion, is not to be taken as necessarily coming from or being directed at a specific philosophical or mathematical position. Gödel's argument is a readily generalizable one adaptable to mathematical systems based on numbers, sets, or computational models of many kinds, just as Sextus distinguishes the general features of skepticism from its "special" part, in which skeptics argue against "each part of so-called philosophy." There is also no inconsistency with skeptical methodology even though Gödel's arguments appear in mathematics, since as we saw before, the contrast between skeptical appearance and dogmatic claims to truth is a formal distinction, not one based on theories of perception or cognitive impressions. If anything, skepticism may be more at home within logic as a formal and nonempirical discipline.

We saw also that the skeptic, in developing isostheneia and the problem of the criterion, does not thereby reject the proposed standard of knowledge but only curtails its claim to absolute truth. This cautious holding back of Pyrrhonian criticism is related to the precise form of the incompleteness results. Both incompleteness theorems, in asserting the unprovability of mathematical propositions, crucially depend on the provisional assumption of the consistency of the system under consideration. If the system is inconsistent, then all propositions are provable in it, so there are no undecidable sentences and consistency can be "proved" in particular. Gödel's theorems are conditional ones whose antecedents include the statement of consistency of the axiom system under view. Formally, Gödel's theorems are Δ_2 sentences in the arithmetic hierarchy, both being theorems of the form $P \rightarrow Q$, with P and Q both Π_1 or universal sentences: for example, "*for all* proofs ***p***, ***p*** is not a proof of the sentence '**0 = 1**' " states consistency. Methodologically, the conditional form of Gödel's theorems makes them part of a critique directed against a potential dogmatic foundation for mathematics. The conditional form cannot be reduced to a lower logical level, so any instance of Gödel's proofs is always conditionalized by the consistency statement for the "candidate" for a mathematical foundation.

This technical proviso is neatly echoed in Pyrrhonian methodology. When the skeptic is questioned about the status of her own skeptical pronouncements, just as Gödel might be questioned about the status of an incorrectly stated *un*conditionalized theorems, the skeptic replies that her conclusions are framed only within the context of a dialectical debate between herself and an antagonist who claims certainty based on some given criterion of truth. The truth of the dogmatic starting point is only conditionally assumed in order to carry to completion the construction of isostheneia, and the formal residue of Gödel's immanent critique of the Hilbert program as a typical dogmatic origin is found in the conditional antecedents of the incompleteness theorems. Mathematical undecidability, like isostheneia, is based on a particular "concession" of dogmatism, and thereby only conditionally refutes an absolute criterion of truth, which is not to show that there is absolutely no criterion of mathematical truth. There is no opposition of principles, but a single immanent critique from within one system and an absence of any genuine positive doctrine. Gödel effectively created Montaigne's "negative language" in which skeptics could express their criticisms without overstating them, and restricted to metamathematics, this is one of Gödel's great achievements: the construction of a complete theoretical apparatus for formulating and proving relative undecidability and unprovability conditions.

It is tempting to try to obviate the conditional dependence of Gödel's theorems on incompleteness and the unprovability of consistency with a kind of metalogical argument: Gödel proved that he could show conditional undecidability if arithmetic is consistent, and if arithmetic is *in*consistent, then everything is provable, including Gödel's theorems. So absolutely either arithmetic is consistent or it is not, and in either case, Gödel's theorems are provable. The point of this metalogical rejoinder would be to mitigate Gödel's antifoundational conclusions, the argument from metalogical *tertium non datur,* or excluded middle, being taken to show that Gödel's conclusions are only as strong as the systems in and against which they are constructed, and that such antifoundationalism is therefore self-refuting or self-annulling. Is the conditional assertion of the unprovability of consistency at least not known, absolutely, even as that assertion means that *everything* may be provable anyway? That account would be contrary to Hume's observation that if the truth of the dogmatic starting point is taken to be thrown into doubt by skeptical argumentation, then such doubt extends to the skeptic and completes the annihilation of reason. The attempt to mitigate Gödel's antifoundationalism is the last step before a parallel annihilation in mathematics, because the metalogical alternative—that in inconsistent mathematics, one can prove anything at all—is in this context incoherent. It represents a Humean condition of reason destroyed in trying to repel a skeptical assault, shown by "proofs" that prove anything and everything—that is, a meaningless, undifferentiated enumeration of symbols. Gödel's theorems, in this way, lead to the same mitigated skepticism we saw before in Lakatos. Gödel's work shows that the dogmatism needed by Hume's schizophrenic-mitigated skeptic is mathematically necessary and the mathematician is forced into such provisional dogmatism by the practical need to carry out research. At the same time, most any assumptions chosen are subject to Gödel's skeptical conclusions, so as far as mathematical foundations are concerned, the mathematician lives out the conditions of Hume's schizophrenic. Or better, as discussed earlier, this mitigated skepticism can be thought of as keeping radical skepticism at bay to achieve pragmatic goals, while acknowledging the force of unmitigated skeptical assault.

Unlayering the skeptical structure implicit in Gödel's work provides a way back into *Proofs.* Lakatos's mitigated skepticism now can be seen as the natural interpretation of Gödel's mathematical critique, which could be carried out within many of Epsilon's Euclidean proof systems. Lakatos, of course, knew that Gödel's results applied to most of Epsilon's formalisms, but he did not notice that these mathematical results are again the expression of the skeptical philosophy he himself embedded in the method of proofs and refutations. Let's then continue the insight that Gödel has formalized several core techniques of Pyrrhonism into a closed interpretive system in which theorems, metatheorems, and meta-metatheorems on relative undecidability and the relative unprovability of consistency are formulated and proved. Gödel, from this perspective, would have achieved the perfection of Pyrrhonism in mathematics, apparently avoiding the critical dangers of the peritrope, or reversal, mentioned before as sometimes deployed against an opponent to show the failure of *his* expression of knowledge to meet an implied or expressly stated criterion of truth.

This skeptical gambit turns out to play a central role, along with the problem of the criterion and isostheneia, in the heuristic structure of Gödel's proofs. The remarkable role for the peritrope in Gödel's theorems is the peritrope's alternative positive use, meaning that the turning of the tables effected by the reversing gambit may show that the speaker of the argument *does* indeed meet the criterion of truth that is used as the speaker's own. For us, the peritrope appears not in the content of either of Gödel's theorems, as do isostheneia and the problem of the criterion, but as the proof of the second incompleteness theorem. Indeed, it is because of this unusual proof that Gödel's second incompleteness theorem is one of the most remarkable corollaries in all mathematics.

The second theorem is proved by a thought experiment in which one reviews the arguments and machinery making the first theorem possible—the use of the Chinese remainder theorem, the elimination of inductive definitions, the syntactic definition of proof, and so on—and then sees that these informal mathematical arguments are all entirely formalizable within the system under consideration, and that they use no mathematics requiring more power than several weak arithmetic operations: that is, they meet just the criteria of formalization that have been informally examined. The next and final step is to realize that the only assumption needed to carry out the proof of the first theorem, in addition to the arithmetic or set-theoretic axioms, is the form of the formalized statement of consistency, the "P" antecedent in the "$P \rightarrow Q$" form of the incompleteness theorems. A weak or misstated version of the first theorem assumes consistency externally, appealing, for example, to the existence of a standard model of arithmetic, perhaps to establish the truth-value of Gödel's undecidable sentence. But the strong version appeals to the thought experiment and, by careful inspection of the first theorem's proof, improves the proof of the second theorem to require only the formal statement of consistency within the system being considered. By this thought experiment, it can be seen that the first theorem itself is provable in, say, *Principia* or Peano Arithmetic, and from this fact, with a few additional lines of elementary argument, the second theorem follows—that consistency implies that consistency is unprovable. The essence of the proof of the second theorem is that the proof formalizes the informal mathematical proof of the first incompleteness theorem; and then that the conditional second theorem is formally provable, too.

There is no self-refutation here, as is typical with applications of the peritrope, but there is a quasi-existential reversal appearing at a precise stage of a sequentially elaborated argument. The first theorem is itself not a formal object, at least initially; it is an informal mathematical proof, through which a claim is established about the limitations of a certain formal model of mathematical reasoning. As this happens, two versions of the first theorem come into play: the informal one proved by Gödel, and the formal representation of that proof, a representation immediately needed in the proof of the second theorem. When, hypothetically, the tables are turned on Gödel and it is "proposed" that the standard of truth that is applied to his object of study become the standard for the propositions he makes, that is when Gödel's first theorem is translated from the informal domain of mathematical argumentation into the formal domain—this peritrope, far from contradicting some needed feature of the first theorem, turns the first theorem into the second. Plato describes the peritropical self-refutation against Protagoras as "this most exquisite feature" of the refutation.[10] Some two thousand years later, in his *Set Theory and the Continuum Hypothesis,* Paul Cohen, who proved the undecidability of Cantor's continuum hypothesis in 1963, remarked of the proof of Gödel's second theorem that it was perhaps the most subtle point in all of mathematical logic.[11] Cohen and Plato perhaps are reacting to similar arguments deployed millennia apart with equal amazement. Only unlike Plato's version of Protagoras, Gödel appears able to dwell within his mathematical skepticism without the pain of self-refutation. His constructive and generalizable method for mathematical isostheneia as well as his mathematical formulation of the problem of the criterion also allow for the iterated self-application of metamathematical concepts following as the coup de grâce. Perhaps living the entire mathematical life requires a kind of mitigated-cum-schizophrenic skepticism, but within the limited scope of Gödel's arguments, there is powerful skeptical coherence and immunity from standard attacks. Gödel *seems* to have perfectly translated Pyrrhonism into metamathematical method.

Or is that too hasty a conclusion? If we could interpret a largely consistent mathematical skepticism even limited to the scope of Gödel's theorems, it would seem that contrary to *Proofs,* metamathematics had evaded history, and that the power of formal systems to absorb informal reasoning was perhaps not absolute, but surely robust. What happens to Lakatos's claim that metamathematics is just more informal mathematics? Is mathematical history dead? As it turns out, Gödel's exquisite technique is subject to the logic of proofs and refutations, and to history as well.

The ingenious peritropical technique by which the second theorem is derived from the first does not automatically preserve *all* the properties one might wish a formalized metamathematical proof to have. What one would hope for initially, and probably expect, is that the second theorem would not essentially depend on the form the translation takes. For example, there are many alternative Gödel numberings possible and a choice among these, or at least a reasonably large class, does not affect the outcome of the proof. Modulo these details, such a proof would be invariant with respect to the choice of translation. This invariance would provide a kind of closure on Gödel's results, which would then be distinguished by their rigorous conceptual and skeptical completeness.

But the second theorem does not have this property, and for mathematically important reasons. When Gödel published his epochal 1931 paper, *On Formally Undecidable Propositions of Principia Mathematica and Related Systems, I,* he intended to elaborate additional details of his proofs in a later work, yet he never did, perhaps due to the reception of Gödel's proofs among mathematicians.[12] This sequel could have contained a study of the exact form of the formalized consistency statement, the crucial antecedent that needs to be made part of the hypotheses of Gödel's results. In the published paper, the explicit form of the consistency statement is relegated to a footnote, inauspiciously introducing into mathematics one of the most curious logical entities ever discovered. The form given for the consistency statement by Gödel is correct. But it is only one reasonable formalization of consistency. While the formalization given by Gödel is precisely the one, up to provable equivalence, needed to make the proof of the second theorem work, there are alternative formalizations of consistency that make the second theorem false; that is, for these choices, consistency *is* internally provable. These constructions are unintended counterexamples to the second theorem, very much Lakatosian monsters, and play the valuable role of forcing a more painstaking analysis of the incompleteness proofs to determine the further conditions, left unspecified by Gödel, that show that his particular choice of canonical representatives for provability and consistency is theoretically justifiable. This analysis was unknown to Gödel, took decades to produce, and is a lemma-incorporating improvement to his original and, nonpejoratively, relatively naive results, his genius and originality notwithstanding.

The core of this theory of formalized proof predicates and consistency statements is a set of conditions examined originally in the context of Gödel's work by David Hilbert and Paul Bernays in the second volume of their *Grundlagen der Mathematik,* published in 1939, eight years after Gödel's paper, though other publications suggest that the idea behind the proof conditions they identified as sufficient for the second theorem was in the air somewhat earlier. While it may originally have been thought that the Hilbert-Bernays conditions, later improved by Martin Löb, merely identified some "obvious" properties that a formalized proof predicate, and by implication a formal consistency statement, should have, the history of Gödel's results from the 1930s on shows, as Lakatos again would say, that Gödel did not prove what he set out to prove in his second theorem. Not that there is a mistake in the second theorem; rather, as in so many cases of the historical development of a mathematical proof, the elaboration of lemmas and conditions needed to make Gödel's proof work eventually led to a novel theoretical treatment of the concepts involved that was not anticipated at the inception of the original result. It is precisely this entry of mathematics into its own history that delivers Gödel and modern formalism from ahistorical mathematical skepticism and back into history. The process occurs also as Lakatos described in his account of the formalization of Poincaré's algebraic proof of Euler's theorem: via a significant heuristic falsification from the informal desiderata for the formal proof, only here it is the very concept of proof that is informal, and which we are trying to formalize.

Already in 1936, it was discovered by Barkley Rosser that desirable technical improvements, specifically the elimination of so-called ω-consistency in favor of simple consistency in the first theorem, appeared to require a special choice for the formalized proof predicate, or the arithmetic formula that enumerates numerical codes for theorems and their proofs. But this would only be the first time that the problem of the "correct" translation procedure into formal metamathematics would be raised, and always motivated by technical and theoretical issues in Gödel's proofs—just the kind of critical mathematical changes discussed in *Proofs* when Gödel's proofs are taken as informal mathematics. Rosser created his alternative formal proof predicate *Prov* from a formalization of "***p*** is a proof of ***s*** and no shorter proof of not-***s*** exists," instead of Gödel's so-called intensionally correct "***p*** is a proof of **s**," denoted by *Bew,* for *beweis,* or "proof." These two formulas are extensionally equivalent, meaning they enumerate exactly the same sets of Gödel numbers of proof sequences and theorems.[13] While Rosser noted that the two proof predicates have very different properties, he did not comment on building a consistency statement using his new proof predicate, but indeed, such an extensionally correct yet intensionally incorrect formalization is provable and a Lakatosian monster for Gödel's second theorem. It is entirely a question of one's informal interpretation of these theorems, too, as it is only *for us* that intensionality matters, even as *this* condition itself is translated into some mathematically useful set of lemmas or conditions.

The problem—which could barely be called a controversy—regarding the correct form of the consistency statement simmered in the mathematical literature for many years. There were criticisms that results related to the second theorem, and the intensional formulation of consistency, were being misstated, although there was also concern over unsolved theoretical difficulties directly caused by intensionality.[14] One can occasionally find misrepresentations of the second theorem in popular accounts of Gödel's work as well as histories of modern logic, the mistake being the incorrect assertion that Gödel proved the impossibility of proving *any* form of consistency of a formal system within the system.[15] This is just false, and a significant theoretical error, not a trivial or merely technical oversight; these are historical falsifications like those provided in *Proofs* that are symptoms of an incorrect understanding of the fallibility of informal proofs.

The translation problem arose again in 1952 when Leon Henkin posed the problem, settled by Martin Löb in 1955, of whether the self-referential mathematical sentence asserting of itself that it *is* provable was indeed provable and therefore true, unlike Gödel's undecidable sentence, which asserts its own unprovability, is not provable, and is thus true. Henkin's problem, Georg Kreisel points out, was ill formed as stated since its resolution could depend on the specific formalization chosen for provability.[16] Taken in isolation from its historical context, Kreisel's objection may seem a bit pedantic, since "everyone knows" the preferred choice of the provability predicate. Nonetheless, that reaction betrays a historical forgetfulness: Gödel had taken logicians over a threshold from philosophy to mathematics, and there were still theoretical issues left unresolved by the choices Gödel had made. In Lakatos's heuristic terms, it is no more than monster-barring. Kreisel's perspective, in addition, was that of a sophisticated proof theorist for whom such distinctions mattered, and these later came to be of importance for mathematical logic.

In any event, these developments mark the transition from Gödel's initial naive conjecture—that consistency is unprovable and this result does not essentially depend on the formalization chosen—to a theoretical, and hence mathematically acceptable, treatment of consistency and proof predicates. The improvement to invariance consists in identifying, as did Kreisel and others, that the implicit choice of formalization is a hidden and false lemma in the original proof. One may say that *Bew,* not the monster *Prov,* is what Gödel surely intended, that *Prov* is a classic concept-stretching monster, and the theorem just is correct as stated. That is, Lakatos's account of concept stretching as the source of unintended refutations can be replayed exactly. Fair enough; but why? *Bew* was to be kept, though only by improving Gödel's conjecture, so that *Bew* appears as a proof-generated concept and *Prov* as Rosser's proof-generated monster. That proof analysis took decades to discover and shows the perfect relevance of Lakatos's heuristic theory to metamathematics itself.

It was not until the late 1950s that Solomon Feferman investigated in some detail the difference between the first and second theorems, and how the truth of the second theorem depends on intensional choices for the consistency statement.[17] Feferman deals directly with the question of whether the "natural," intensionally correct definition of formal consistency, *Con,* can be theoretically justified as the sole choice, and he essentially answers this question in the negative. The second incompleteness theorem, unlike the first, relies on choices that cannot be completely justified within Gödel's original epistemological framework, which, I presume, does not entail intensionality—even the Stoics did not claim that. Feferman shows that there are general conditions, not merely isolated counterexamples—implicitly answering perhaps a wishful Cauchy-style exception-barring with no further explanation—under which consistency *is* provable using an extensionally correct, but from the retrospective standpoint of Gödel's results, "intensionally deviant" representation of formal provability:

> Rather than contradicting Gödel's second underivability theorem [these results] show the importance of a precise method of dealing with consistency statements, at any rate for theories with infinitely many axioms [such as Peano Arithmetic]. A first reaction following such realizations might be to restrict attention to a certain class of "natural" formulas α in problems of arithmeticization. . . . However, we shall obtain . . . results through the use of arbitrary formulas α which should be of interest even to those who would otherwise thus restrict attention. There is nothing "wrong" with the use of arbitrary formulas α; rather, the guiding consideration should be to investigate how different restrictions on the *choice* of α affect the results by arithmeticization.[18]

Feferman then uses an ordering on representations of axioms to provide a framework for choosing a single arithmetic sentence, up to provable equivalence, to express consistency, the "natural" choice being the minimum in this ordering. But for arbitrary—that is, those with an infinite number of axioms—arithmetic systems, he shows that no such minimum or maximum in the ordering exists. As Feferman writes,

> The moral of these theorems is not to reject the use of particular numerations for known particular axiom systems ***A***; for example we still consider "natural" the definition-of ***P***[eano Arithmetic] as a finite set of axioms and axiom schemata. Rather, it is to reject the use, as a well-defined idea, of sentences *Con(**A**)* associated with arbitrary systems ***A*** . . . [T]here is no natural or favored description . . . of an extensionally given infinite set of axioms.[19]

Under fairly unrestrictive conditions, there is then no natural choice for expressing consistency. What is "natural" is a matter of historical choice made against a background of mathematical tradition adapting to a radical new set of ideas. Another way of putting Feferman's point is this: if one performs the thought experiment of carrying out the proof of Gödel's first theorem using only extensional notions of provability—notions depending only on the numbers formally representing provable sentences, but not on how these numbers are variously denoted—the pathological consistency statements can "appear" completely normal and the second theorem need not hold; it really is falsified. It is only when one introduces the needed intensional conditions taken from additional informal mathematical arguments that Gödel's thought experiment works the way it should. Gödel's second theorem is "provably" a piece of informal mathematics, exactly as Lakatos casually observes in the opening pages of *Proofs.*

After Feferman, the theoretical characterization of *Prov, Bew,* and *Con* took another turn, just as it took several attempts to reformulate the theory of the integral in nineteenth-century mathematics. In addition to Feferman's approach, there is now a well-developed theory based on the semantics of modal-logical systems, created originally by Saul Kripke, that provides further mathematical justification for the choice of *Bew* over *Prov,* and *Con* over a monster *Con** constructed from *Prov* rather than *Bew.* The Hilbert-Bernays-Löb conditions are expressed using a modal operator "$\Box \boldsymbol{p}$" with the intended interpretation that "it is provable that $\boldsymbol{p}$" and implications drawn for options for translating "$\Box$" as provability using the methods of modal logic.[20] For example, using the modal representation of proof conditions, it can be shown why the Rosser undecidable sentence cannot be replaced by one using the canonical proof predicate to eliminate ω-consistency, and how to distinguish *Bew* and *Prov* by critical properties, among the most important being that the coextensiveness of *Bew* and *Prov* is *not* provable in Peano Arithmetic; that is, Peano Arithmetic does not prove that *Prov* can be substituted for *Bew.* These results provide a non–ad hoc theoretical differentiation between *Bew* and *Prov* not available to either Gödel or Rosser, and finally explain the pathological status of *Con**. It is results such as these that justify the crucial qualification of the unprovability of consistency, that as George Boolos says in his *The Unprovability of Consistency,* "Although *Prov* is coextensive with *Bew,* $-$*Prov* ('$0 = 1$') does not express the consistency of *PA* [Peano Arithmetic]; at best it merely asserts that every proof of '$0 = 1$' that there might be occurs later than some proof of '$0 = 1$.' "[21] On the other hand, Boolos neglects the problem-background, as well as the heuristically sound reasons for privileging *Bew* over *Prov,* which explain what the essentialist and monster-barring "correctly expresses" means in this context. Essentialism is fine as a heuristic strategy, but that is all.

What is important is that these characterizations are mathematically, rather than philosophically, motivated and justified. Mathematical history, therefore, leads out of the stagnant quagmire of pure skepticism through a recognition of the specific historical development that led to the current understanding of incompleteness. An ahistorical philosophy of mathematics will not be able to explain the interaction between philosophy and mathematics that took place in Gödel's work and its aftermath. Ahistoricism is not philosophically inconsistent, even though it will lead those trapped by it into dogmatic conceptions of "naturalness" and "canonicity." More broadly, if the second theorem crucially depends on technical choices for the form of *Con,* then the sometimes romantic philosophical implications of the second theorem on the limitations of the human mind, or whatever, are masked behind a specially chosen set of mathematical conditions that are not generally justifiable, let alone so within Gödel's implicit mathematical skepticism. The Pyrrhonian "immanent" interpretation of Gödelian epistemology reveals how arbitrary the invocation of an intensionally correct representation of provability is; it is not to be found in a dogmatic setup, whether that be Hilbert's finitism or other special philosophical starting points. There is nothing in the epistemological stage setting for Gödel's proofs that allows such a move. The philosopher Gödel needs to argue like a true Pyrrhonian skeptic, and he can to a certain extent; this is the basis of his mathematical criticism. But he is forced to shift gears and compromise his position as his proofs develop—over the years!

What is illustrated here is a historical conflict between mathematical progress and philosophical origins. The "deviancy" of Rosser sentences is a mathematical result that does not quite mesh with the elementary skeptical epistemology needed to work through the incompleteness theorems. Nevertheless, it is heuristically unsound to replace an interesting nexus of theoretical difficulties with apparently ad hoc prescriptions that cover up good problems.[22] Gödel changed a philosophical problem even as he improved it into a mathematical one. He also did not solve the problem he set out to solve. Ahistorical mathematical philosophy does not recognize this mathematically essential fact. It is not that consistency might really be mathematically provable but that the informal notion of consistency was too naive to withstand the theoretical development that Gödel initiated. To recognize this historical event is to diminish the force of Gödel's work against the old foundationalists, yet that at least acknowledges that Gödel began a new branch of mathematics by partially jettisoning an old set of philosophical standards. I agree that *Bew* "better expresses" provability than Rosser's *Prov* —but because it makes a mathematical theory work in a desirable way, not because of a nonexistent accurate translation.

Therefore, it is just Lakatos's problem of translation between informal and formal mathematics that is at the heart of an intriguing tale in the history of twentieth-century mathematics. The historical process shows that the second theorem lives just on the border between informal and formal mathematics. If one ignores the translation that takes place here between these domains, there is no exclusion of consistency monsters. On the other hand, bringing in the Hilbert-Bernays-Löb conditions changes the original conditions for carrying out Gödel's proofs: from "working inside" a given formalism, to working inside the formalism with some additional choices made after certain results are proved about the formalism and reflected on with heuristic as well as historical hindsight. The justification is perfectly sound, let there be no doubt; still, the issue is how that justification is introduced into a mathematical argument and how the argument then is changed. The process demonstrates that, at least here, mathematics cannot be completely identified with its formal, metamathematical representation, and is also on the border between the historical and ahistorical, the temporal and timeless. That is what Gödel's exquisite proof proves.

5 | Reason Inverted

The intriguing idea that many leading scientists and mathematicians have, purposely or not, managed to conceal the heuristic logic underlying their discoveries dates back at least to Archimedes. The seventeenth-century mathematician John Wallis remarked of Archimedes's geometric treatises that it is

> as it were of set purpose to have covered up the traces of his investigation, as if he had grudged posterity the secret of his method of inquiry, while he wished to extort from them assent to his results. . . . [And] not only Archimedes but nearly all the ancients so hid from posterity their method of Analysis (though it is clear that they had one) that more modern mathematicians found it easier to invent a new Analysis than to seek out the old.[1]

It is not as if these methods would then become known, however: the ancient method of analysis to which Wallis refers has also been considered by some to again be a methodological "secret" during the period of early modern physics from Galileo to Newton, with even Galileo and Descartes noting the hidden methods of the ancients.[2]

The method of analysis referred to here, or analysis-synthesis, is that of Pappus of Alexandria, who lived around the end of the third century C.E. Pappus described this brief but lucid heuristic within a large set of treatises in which the method takes on a specialized role for developing geometric results only. In the general terms in which it has influenced mathematicians and scientists, the method of analysis-synthesis is closely related to the modern one of proof by contradiction. Roughly: Given a conjecture, derive consequences from it. If a consequence known to be false is arrived at, the conjecture was false, and by reversing the order of the derivation, a proof of the negation of the original conjecture follows. If a consequence known to be true is reached, reverse the order, if possible, and if the conjecture can be derived from this true consequence, then it too is true.[3] Typically, if the goal is to prove conjecture C, assume not-C and derive a contradiction, if possible. The negation of the contradiction is tautologically true, and hence, the chain can be reversed to prove C as desired. That is not what Pappus said; in practice, however, that is one of the easiest generic uses of the method. The power of the method in that form is due to the ability now to search for a proof of *any* contradiction, or even C itself, since (not-$C \rightarrow C) \rightarrow C$ and (not-$C \rightarrow \phi) \rightarrow C$ are all tautologies, where ϕ is any contradiction: deriving C or any of an infinite set of contradictions from not-C provides a proof of C.

What is relevant here is the reversal between the order of discovery and then justification. When a proof is created by reversing a long chain of inferences leading to a contradiction, what is presented as "first" in the logical order may be what was discovered "last," so that the logical justification can end up concealing the heuristic path that led to it. Aristotle portrayed this reversal centuries earlier than Pappus in discussing an analogous relationship between ends and the means to achieve them:

> Having set the end they consider how and by what means it is to be attained; and if it seems to be produced by several means they consider by which it is most easily and best produced, while if it is achieved by one only they consider how it will be achieved by this and by what means this will be achieved, till they come to the first cause, which in the order of discovery is last. For the person who deliberates seems to inquire and analyze in the way described as though he were analyzing a geometrical construction . . . and what is last in the order of analysis seems to be first in the order of becoming. And if we come on an impossibility, we give up the search . . . but if a thing appears possible we try to do it.[4]

Among its several roles, Lakatos presents the method of proofs and refutations as an improvement on Pappus's method. *Proofs* also contains much discussion on "starting points" for discovering proofs, including inductive guessing, such as enumerating the "Euler characteristic" $f(V, E, F) = V - E + F$ for various polyhedra and looking for a "law"; and deductive guessing, through which a proof idea, such as cutting and pasting together polygons and polyhedra to form new polygons and polyhedra, can be used to infer changes to $f(V, E, F)$, and hence infer for which classes of polyhedra does $V - E + F = 2$. In such a case, it can easily be that the conditions discovered that finally justify a theorem and its proof give no clue to explaining *why* the proof proves. Either inductive or deductive guessing may need to be reversed to offer a proof in which the route to discovery disappears.

In *Proofs* it is Epsilon who presents proofs in traditional "Euclidean-deductive" style, meaning to start with some so-called primitives taken to be understood, a set of defined terms, and a traditional deductive proof consisting of lemmas leading to a theorem. Our school proofs in geometry are of this type, but a great deal of published mathematics is presented in this way as well. For Lakatos, "deductivist style"

> starts with a painstakingly stated list of *axioms, lemmas* and/or *definitions.* The axioms and definitions frequently look artificial and mystifyingly complicated. One is never told how these complications arose. The list of axioms and definitions is followed by the carefully worded *theorems.* These are loaded with heavy-going conditions; it seems impossible that anyone should ever have guessed them. The theorem is followed by the *proof.* (*P&R* 142)

Lakatos gives several examples from well-known mathematics texts to make his point that Euclidean style is often obscurantist, and positively excludes a rationale for definitions and needed lemmas. Little in Lakatos's description is controversial. What is key is that—and how—Lakatos explains the opaqueness of Euclidean-deductivist style as a philosophical problem. Deductivist style is antithetical to the entire fallibilist spirit of *Proofs* in that it assumes that the growth of mathematical knowledge is driven by increasing truth instead of reducing falsity. The difference is that the former eliminates the generation of concepts and lemma-incorporation of counterexamples that constitutes the relevant heuristic guidance for why the deductivist proof works and how anyone would have come up with the elaborate machinery presented at the proof's start. Pappus's method of analysis-synthesis provides one model by which such inversions of conceptual priority can take place, and this particular concealment of discovery is attested to by the complaints of Wallis and others.

In the method of proofs and refutations, the potential for concealment is significantly increased because of the complicated linkages, possibly taking place over decades, used to reformulate concepts and supporting lemmas, and finally making a proof work: ideas may be borrowed from different fields of mathematics altogether, errors may be corrected from earlier generations having far fewer mathematical resources, and naive approaches to conjectures may have in the meantime matured. For example, we saw that Cantor's development of his theory of infinite ordinal and cardinal numbers out of his work on point sets and trigonometric series has heuristic roots reaching back to Dirichlet's proof of Fourier series convergence some fifty years earlier; the heuristic background clearly shows why transfinite ordinals should have occurred to anyone at all. At the same time, heuristics do not constitute justification. Complete definitions and an orderly presentation make an informal proof rigorous, and such justification may still fail, as happened repeatedly during the nineteenth century. It is not that Euclidean proofs are pointless, though they may have only one point, to justify a theorem as true. For Lakatos, this truth is often created or discovered via the method of proofs and refutations, or other heuristic methods, and that unity of the logics of discovery and justification is, as he says, the "most important aspect of the method of lemma-incorporation" (*P&R* 37). It is also not as if Euclidean proofs cannot get by without their heuristic past; they can and do. But heuristic reduces falsity and creates growth and progress. If those are the goals, they are served poorly by Euclidean style. The aim of rigorous logical justification, as an end, can be mistaken for explanation, and can positively distort explanation by inverting the order of discovery and excluding changes made over time to get the proof to work. The consequences, in Lakatos's mind, are serious and real:

> As long as a counterexample was a blemish not only to a theorem but the mathematician who advocated it, as long as there were only proofs or nonproofs, but no sound proofs with weak spots, mathematical criticism was barred. It was the infallibilist philosophical background of Euclidean method that bred the authoritarian traditional patterns in mathematics, that prevented publication and discussion of conjectures, that made impossible the rise of mathematical criticism. . . . [T]he Euclidean method can, in certain crucial problem situations, have deleterious effects on the development of mathematics. Most of these problem situations occur in growing mathematical theories, where growing concepts are the vehicles of progress, where the most exciting developments come from exploring the boundary regions of concepts, from stretching them, and from differentiating formerly undifferentiated concepts. In these growing theories intuition is inexperience, it stumbles and errs. There is no theory which has not passed through such a period of growth; moreover, this period is the most exciting from the historical point of view and should be the most important from the teaching point of view. These periods cannot be properly understood without understanding the method of proofs and refutations, without adopting a fallibilist approach. (*P&R* 139–40)

Euclidean method is realistically described as the antithesis of fallibilism: given certain assumptions and rules of proof, the conclusion follows, we believe, with certainty. But if mathematical progress depends on the heuristic patterns articulated in *Proofs,* which in turn rely on a philosophy of learning through error in informal mathematics, then it is natural that the method seen as closest to infallible inference is normatively undesirable as either a method of explanation and teaching or a philosophy of mathematical growth. When a Euclidean style is adopted generally, and not just the goal of rigorously justifying a proof, then Lakatos maintains,

> Mathematics is presented as an ever-increasing set of eternal, immutable truths. Counterexamples, refutations, criticism cannot possibly enter. An authoritarian air is secured for the subject by beginning with disguised monster-barring and proof-generated definitions and with the fully-fledged theorem, and by suppressing the primitive conjecture, the refutations, and the criticism of the proof. . . . The whole story vanishes, the successive tentative formulations of the theorem in the course of the proof-procedure are doomed to oblivion while the end result is exalted into sacred infallibility. (*P&R* 142)

The single-minded quest for justification becomes the means by which proof heuristics and their history are concealed and rationality is inverted into its opposite. What gets substituted to motivate the proof, according to Lakatos, is a kind of authoritarianism, and that is no polemical façon de parler. When Epsilon plays his role in *Proofs,* Lakatos has him speak in terms of the "intuitions" needed to begin one of his proofs, akin to what is often expected as assumed background knowledge in mathematics texts. These intuitions really are meant somehow to be inarticulately grasped; they have no further explanations, or very little explanation, and are just ahistorically and ineffably "given." Hence, the starting point could be: "all polyhedra have faces"; "all faces have edges"; and "all edges have vertices," with faces, edges, and vertices being taken as primitives. Or for an example from the account of Gödel's theorems, presented earlier, it might be "the" sentence *Con* that "correctly" expresses the consistency of arithmetic, and the formalized Hilbert-Bernays-Löb conditions. What matters for the reading of *Proofs* here is not whether a complete Euclidean program really can be carried out, but the associated philosophical and pedagogical consequences. For Lakatos, the primary consequence is that the Euclidean ends up promoting dubious conceptions of intellectual intuitions, insight, or attitude needed to get a Euclidean proof started: in effect, the residue of ineffable genius or talent. Thus, Epsilon is given to explain "intuitions" by ventriloquizing a definition provided in the history of *Proofs* by Descartes: that intuitions are "nondubious apprehensions of a pure and attentive mind which are born in the sole light of reason."[5] Less philosophically, and quite commonly as Lakatos points out,

> Some textbooks claim that they do not expect the reader to have any previous knowledge, only a certain mathematical maturity. This frequently means that they expect the reader to be *endowed by nature* with the "*ability*" to take a Euclidean argument without any unnatural interest in the problem-background, in the heuristic behind the argument. . . . [Given this ahistorical attitude, research] will be understood only *by the few* who actually know the problem-situation. This game can only be played—not always with success—by and for *a selected guild of experts.* While claiming objectivity, it *fosters a private guild-language,* atomizes science, suffocates criticism, makes science authoritarian. (*P&R* 142n1, emphasis added)

The sequence, then, is: reversal of heuristic discovery → Euclidean style → loss of heuristic and historical logic of discovery → need for substitute intellectual intuitions or their equivalent. The end point here of inexplicable "abilities" and elite experts is, not surprisingly by now, an application of just the type of intellectual intuition against which Hegel framed his pedagogically motivated *Phenomenology.* We saw at the start of this account of *Proofs* and the *Phenomenology* how Schelling's esoteric philosophy was just what Hegel wanted to prevent and perceived as the worst enemy of constructive Bildung. Quite similarly, standard Euclidean style is opposed to Lakatos's historical philosophy and supports itself with the same kind of spurious epistemology ostracized by Hegel in his preface. Euclidean style, which can be distinguished from the goal of rigor, is according to this view elitist, authoritarian, and irrational when those words are given the meanings they have for either Hegel or Lakatos. Even in terms of content, Schelling's intellectual intuition is barely improved on, if at all, by its mathematical variants. Mathematical intuition has not been just tacitly or casually adopted as an informal epistemology. For instance, it was actively promoted by no less than Gödel through his mathematical Platonism, expressed through his experience that the axioms of set theory "force themselves upon us as being true."[6] Gödel's "perceptions," though, were not entirely consistent. Lakatos notes that Gödel originally thought that his axiom of constructibility, or $V = L$, used to establish the consistency, but not the independence, of the axiom of choice and Cantor's continuum hypothesis in 1938, was "a natural completion of the axioms of set theory," but changed his mind later as its awkward consequences unfolded.[7] Why did Gödel's perception change? What theory of perceptual error explains that? In fact, he had no such perceptions in any substantive sense. Gödel was a genius who convinced some of his personal Platonism perhaps through his authority, and likely not because others were "perceiving" with him.

Another side of the problem of elites and intuitions, and the need for history as its cure, is that mathematical history, according to Lakatos, has been regularly falsified to maintain an infallibilist ideology. That is, when errors occur, it is rare for them to be acknowledged as real, since mathematics is not supposed to be conjectural and contain significant mistakes, and *Proofs* provides a stream of examples in Lakatos's commentary of such falsifications in nineteenth-century mathematics. So, for instance, Lakatos claims that "antihistorical historians attribute our modern general concept of a real function to Dirichlet" (*P&R* 151), when Dirichlet was at best ambiguous about whether his novel characteristic function of the rational numbers, called $\boldsymbol{Q}(x)$ above, was a legitimate mathematical function; or the mathematician George Hardy, the Bourbaki, and the *Encyclopädie der Mathematischen Wissenschaften* all erroneously attributed to Abel the discovery of uniform convergence.[8] On the discovery of various hidden lemmas of Cauchy's proof of Euler's theorem, Lakatos writes that "when it is first discovered, the hidden lemma is considered an error. . . . Historians however cannot imagine that great mathematicians should make such errors. A veritable programme of how to falsify history" (*P&R* 46n1) can be found in Poincaré's *Science et Méthode,* in which he explains that geometers before 1820 did not make mistakes, that if their words were taken too literally one would have to conclude there was no mathematics before 1820, that they really intended the correct proofs and just failed to write them down in their entirety. As Lakatos observes, "Nothing is more characteristic of a dogmatist epistemology than its theory of error," Poincaré's theory being his account of what was "truly known" but "carelessly omitted" (*P&R* 31n3). Another tactic is to distinguish so-called technical errors from conceptual innovations, and to assimilate the latter to the former. Falsification of mathematical history, therefore, supports antihistorical pedagogy by concealing the errors and fallibilist origins in the scientific archive. As such, infallibilist conceptions of mathematics, theorems, and proofs are created in our culture as mock truth. Even "axiom" before Euclid's time stood for statements put forward to be examined in dialectical debate for their consequences, but *without* admitting their truth. It is one of the ironies of history that Euclidean geometry provided absolutely true and infallible knowledge until the discovery of non-Euclidean geometries, while today, with the cornucopia of axiomatic approaches to set theory, the meaning has just about come full circle to the original one. But in mathematical education and cultural attitudes, Euclidean dogmatics are still often promoted.

The problems with Euclidean style, then, are not theoretical ones of whether mathematical intuition really exists. There is such a thing as talent, and there are geniuses with intuitions, such as Srinivasa Ramanujan, who conjectured and proved novel theorems of advanced analysis by reading British schoolbooks in colonial India, or even Gödel, the most recent exponent of mathematical perception. The question is whether it is normatively desirable and historically accurate to base a conception of the growth of mathematical knowledge, its transmission within and across generations, and our self-understanding as historical beings on the idea of "intuitive grasp," as Hegel saw in Schelling; or does one actively contribute to the historical development of mathematics and culture through a pedagogically inspired philosophy that explains the mathematical present through its, or rather our, past. These are cultural and societal problems of practical pedagogy, including that of history as Bildung, cultural formation, and self-education. The modern analogue of Schelling's intellectual intuitions is exposed by Lakatos as excluding change, growth, and progress in mathematical knowledge. Such is Lakatos's criticism of the role of infallibility and mathematical ideology in contemporary culture.

Lakatos's antidote to Euclidean-deductivist style, then, includes heuristic pedagogy, *plus* classical Euclidean justification as one goal of mathematical proof, *plus* historiography acknowledging the conjectural status of proofs as well as distinguishing contemporary concepts and standards of rigor from those of the past. Such an antidote is not simple, as it implies a generalized analysis-synthesis in which *two* proofs are taught and learned: the historical-heuristic and the justifying Euclidean proofs, and some relevant historical scaffolding as well. All that is a means for addressing the challenge of explaining and learning the inversions of logical and heuristic order: "One should not forget that while proof-analysis *concludes* with a theorem, the Euclidean proof *starts* with it. In the Euclidean methodology there are no conjectures, only theorems" (*P&R* 107).

History and Language

The historiographical style of *Proofs* is one means of elaborating that Aristotelian reversal. It is, in particular, an articulated means. To return to the role of language and the exoteric expression of rationality in Hegel, we start with a statement of how the biogenetic bildungsroman form of *Proofs* can be used to address the ancient problem of combining the order of discovery with that of presentation. This account is provided by Marx in his once-unknown rough draft for *Capital,* the *Grundrisse.* There, Marx offered a concise description of the biogenetic narrative for his economic theory with its telescoping and reconstruction of past economic categories—much as depicted by Hegel in the preface to the *Phenomenology*—but also with a recognition of the differences between historical order and that implied by the logic of contemporary economic processes:

> Bourgeois society is the most developed and the most complex historic organization of production. The categories which express its relations, the comprehension of its structure, thereby also allows insights into the structure and the relations of production of all the vanished social formations out of whose ruins and elements it built itself up, whose partly still unconquered remnants are carried along within it, whose mere nuances have developed explicit significance within it, etc. Human anatomy contains a key to the anatomy of the ape, . . . [which] can be understood only after the higher development is already known. . . . One can understand tribute, tithe, etc., if one is acquainted with ground rent. But one must not identify them. . . . They [the earlier periods] can contain them [economic relations] in a . . . stunted or caricatured form. . . . In all forms of society there is one specific kind of production which predominates over the rest, whose relations thus assign rank and influence to the others. . . . It would therefore be unfeasible and wrong to let the economic categories follow one another in the same sequence as that in which they were historically decisive. Their sequence is determined, rather, by their relation to one another in modern bourgeois society, *which is precisely the opposite of that which seems to be their natural order or which corresponds to historical development.* The point is not the historic position of the economic relations in the succession of different forms of society. . . . Rather, their order within modern bourgeois society.[9]

The role played by "modern bourgeois society" in Lakatos's biogenetic reconstruction is that of a modern proof, whose proof ancestors appear in "stunted or caricatured form" in Lakatos's reconstruction and mathematical history itself. The "order" presented in *Proofs* is not their order according to contemporary standards of rigor and mathematical truth but a heuristic one leading to modern logical justification.

In both Marx and Lakatos there is a conversion of historical order, adopted from Hegel, used to provide their accounts of the economic and mathematical present, respectively; this borrowed historiographical style is not too explicit, and Hegel is to be kept safely at arm's length. The draft of *Capital,* Marx wrote to Engels in 1861, "is becoming much more popular and *the method is much more hidden,*" and in a letter to Ferdinand Lassalle on the latter's attempt to reconstruct Heraclitus's philosophy through extant fragments, Marx remarked that

> the difficulties you had to overcome in your work I wrote on in a much easier philosopher, Epicurus—namely the presentation of the total system on the basis of fragments. . . . Even with philosophers who gave their work a systematic form, e.g., Spinoza, the real inner structure of his system is, after all, wholly different from the form in which he consciously presented it.[10]

That, too, would be the case with Marx himself. As the translator of the *Grundrisse,* Martin Nicolaus, points out, "The *inner* structure [of *Capital*] is *identical* in the main lines to the *Grundrisse,* except that in the *Grundrisse* the structure lies on the surface, like a scaffolding, while in *Capital* the method is built in."[11]

This structure, including the fundamental role for alienation taken over from Hegel, was also the hidden content found by Lukács before the discovery of Marx's early unpublished manuscripts.[12] We have seen such a hidden biogenetic structure, too, in Lakatos, who makes the problem of hidden rationality one of his explicit concerns. The problem of concealed method with regard to Galileo and Descartes, as mentioned earlier, is now repeated vis-à-vis Marx and Lakatos's covert Hegelianism. In spite of Marx's efforts to make clear the basis for his reconstruction of the economic past, the hidden logic of Hegelianism in Marx is nonetheless veiled, as it is in Lakatos, the concerns of the latter about concealed heuristics notwithstanding. Such are the ironic joys of history.

Now it is necessary to regain our balance. Lakatos explains how to reveal the concealed proof ancestors of a theorem proved in standard Euclidean style, and the method of proofs and refutations builds on the ancient one of analysis-synthesis, thought by some to be a secret method of proof discovery. Lakatos's account is also historical, and that is his great novelty, just as Marx historicized political economy by applying a similar biogenetic historiographical technique to economic history, among other ideas. Marx, too, saw his method as a means of coordinating two narrative logics. Then, just when the daylight of reason almost appears in Lakatos, through the historical means for uncovering the complexities of justification and concealed discovery, Marx is seen hiding one debt to Hegel and a corresponding debt is hidden in Lakatos. One secret method is revealed only to lead to another. Marx's motivation in keeping philosophy out of his social and economic theory was largely political, and even Lakatos, recognizing not just the general low esteem for Hegel in England but Popper's unqualified disdain for Hegel, may also have had a straightforward rationale in concealing his intellectual roots to keep his work, as Marx said, "popular." But Marx was not reforming a contemporary understanding of justification and discovery, he was just coping with politics, while Lakatos is preeminently a pedagogical philosopher. What should be made, then, of Lakatos's apparent Hegelian reason in disguise, not just against his stated intentions of articulating reason in history but his anti-intuitionist polemics?

One lesson here is this very exercise to reconstruct a relevant history behind *Proofs.* That history shows that even the most enlightened and critical historicist reason will not necessarily explain everything, and it is never clear whether the whole account of a proof, philosophical text, or any other historical product has been uncovered. There is also some subversion here of Lakatos's method: while he criticizes deductivist style for mangling and concealing the true history of ideas, he apparently has left his own method as a small historical secret. Reason, contrary to Lakatos's valid intentions in his critique of Euclideanism, may sometimes be effectively deployed without complete clarity and transparency. It is as if a skeptic has noticed a latent dogmatism in exposing the logic of discovery and wants to offer a reminder that *that* revelation too is still conjectural. The precedents of the mathematical history of *Proofs* are partially hidden, to be discovered themselves.

What helps make that possible? How can so much have been discovered? None of Lakatos's philosophical interest in concealment and discovery is expressed in Marx, even as Marx captured how historical narrative can be guided backward by the logic of contemporary economic categories. It is not only motivation that distances Lakatos from Marx. The latter's history is of social relations and labor and market practices, not of ideas or language. In a clear return to Hegel, Lakatos's history is of theories, sentences, conjectures, and mathematical language generally. A history of mathematical language, not social relations and their embodiment in economic processes, provides the foundation for Lakatos's entire approach. Though there are salient similarities between his and Marx's histories, including the use of a biogenetic narrative to characterize the present as history, they differ categorically. It was ostensibly from Popper that Lakatos came to describe what he needed from the philosophy of language as the "objective knowledge" associated with Popper's so-called "third world" of language and its products, primarily scientific theories, but also politics and the languages of government and debate, constitutions, laws, and other forms of social discourse and social consciousness.[13] Popper's "first world" is the physical one and the "second world" is that of mental events, and his three worlds interact in fairly pedestrian ways. He took these three worlds to be largely irreducible to one another as a means of guaranteeing the relative autonomy and objectivity of third world knowledge, although at the unfortunate expense of introducing a heavy-handed ontological distinction to justify a reasonable empirical claim about the autonomy of some linguistic practices. Lakatos's history in *Proofs* requires being able to meaningfully separate linguistic products from their producers for some purposes, just as occurs within mathematics and science proper. Given both the pedagogical and historicist roles for *Proofs,* language is the critical medium for mathematical objectivity, conceptual change, historical reconstruction, and self-formative Bildung. It is also the means by which Lakatos effectively conceals a Hegelian historiography even as he uses it to polemicize against the dangers of concealed heuristics.

Popper's "epistemology without a knowing subject," depending as it does on the idea of objectified knowledge, provides yet another serendipitous confluence of like ideas. That language supplies the essential medium for Bildung and human Geist is one of Hegel's fundamental assumptions about human beings, and especially opposed to Schelling's nondiscursive intuitions. It was even Hegel's fault, from Marx's perspective, that he saw so much of society and culture as expressed discursively. The *Phenomenology* is replete with discussions on language as exteriorized and objectified humanity, or Popperian objective knowledge. Hegel carefully distinguished meaning from its "physiognomic expression" (*PhS* 185), that is, the use of Popperian first world, physical, possibly bodily, symbolic media, and much of the *Phenomenology* is spent discussing political institutions as speech acts, such as their evolution from naive to more sophisticated institutions, passing through forms such as the "chatter" of the Enlightenment, similar to Lakatos's concern with progress made through improved theoretical languages. When Hegel mentioned "daylight" or "sunshine" in the *Phenomenology,* he meant the world of culture as existing only through language and "exteriorizing" our "second nature."

Culture is made possible by language in that many social institutions are characteristically linguistic, much in the spirit of the later Ludwig Wittgenstein and the *Philosophical Investigations.* For Hegel, cultural processes, like those associated with formal laws, though grounded in the language and customs of an ongoing community, also become autonomous from the individuals giving them life:

> This alienation takes place solely in language, which here appears in its characteristic significance. In the world of ethical order, in law and command, and in the actual world, in counsel only, language has the essence for its content and is the form of that content; but here it has for its content the form itself, the form which language itself is, and is authoritative as language. It is the power of speech, as that which performs what has to be performed. For it is the real existence of the pure self as self; in speech, self-consciousness, qua independent separate individuality comes as such into existence, so that it exists for others. Otherwise the "I," this pure "I," is nonexistent, is not here. . . . Language . . . contains it in its purity, it alone expresses the "I" itself. This real existence of the "I" is, qua real existence, an objectivity which has in it the true nature of the "I." (*PhS* 308)[14]

For Hegel, virtually all aspects of culture are externalized and alienated through language from the individuals who created them. Marx, in contrast would provide the theory of alienation and objectification par excellence, definitely *not* of ideas-in-language but of the social relations engendered by modern categories of labor, commodities, money, and capital. Marx offers innumerable insights into objectification and the strange roles played by alienated social products in modern capitalist culture, while Hegel's language ontology is the one needed by Lakatos in his history of objectified and semialienated mathematical theories. When Lakatos, in a brief passage on language and the "personification" of mathematics (*P&R* 146), refers to alienation instead of objectification, he is close to making the same conflation as the "young" Lukács of the 1920s in *History and Class Consciousness,* between a normatively undesirable alienation and an ethically neutral objectification. As Lukács self-critically wrote some forty years later,

> Objectification is indeed a phenomena that cannot be eliminated from human life in society. If we bear in mind that every externalization of an object in practice (and hence, too, in work) is an objectification, that every human expression including speech objectifies human thoughts and feelings, then it is clear that we are dealing with a universal mode of commerce between men. And insofar as this is the case, objectification

> is a neutral phenomenon; the true is as much an objectification as the false, liberation as much as enslavement. Only when the objectified forms in society acquire functions that bring the essence of man into conflict with his existence, only when man's nature is subjugated, deformed and crippled can we speak of an objective societal condition of alienation and, as an inexorable consequence, of all the subjective marks of an internal alienation. (H&CC xxiv)[15]

Suffice it to say, the idea of objectified, "third world" discourse, and its problems, is much older than Popper's philosophy, and there is little apart from obscure terminology to be amended in Hegel's philosophy of language to guard against Popper or others' unneeded ontologies. Feyerabend himself gave a materialist and Wittgensteinian account of linguistic objectivity intended expressly to naturalize Popper for his application in *Proofs*.[16] Those who might agonize over associations of Lakatos with a "metaphysical" Hegel therefore have no ground on which to stand, since it is just language that supplies Lakatos's historiographical and even skeptical ontology, remembering especially the topic-neutral and not "perceptual" interpretation of skepticism and the *Phenomenology* uncovered here earlier. Language for Lakatos is where rationality and reason in history reside, not in the inarticulate intuitions of Euclideans, Schelling, or other intellectual aristocrats. It is a wholly Hegelian view, concealing much in its own expression.

That completes this account of the implicit Hegelian heritage of *Proofs*, and even part of a depiction of its own dissembling. While this led finally to a Hegelian conception of language as underlying Lakatos's historiography, the latter prompts further questions about creating historical narratives in mathematics or science generally. How skeptically, for example, should reconstructions of the past be viewed? Even though Lakatos's history may be understood as one of mathematical theories, how should such mixed philosophical-historical claims be evaluated in general? Should a mitigated skepticism about *their* truths be adopted? Or is another Humean-near-the-abyss skepticism necessary before stepping back to that practical choice? How might all that be unraveled as a problem of historical criticism in itself? Those issues and more are buried within Lakatos's philosophy of science: his methodology of scientific research programmes.

11 | A Changing Logic of Scientific Discovery

It is a book which "saves the critic the trouble of judgment." And why? Because it holds no mystery: "it judges itself." And not only does it judge itself, it also represents itself—together with the story it purposes to represent.

—FRIEDRICH SCHLEGEL on *Wilhelm Meister*

6 | Kuhn, Popper, Feyerabend, Lakatos

The philosophy of science debates of the 1960s and 1970s, involving Kuhn, Popper, Feyerabend, and Lakatos at their center, broke the dominance of logical positivism in Anglo-American philosophy. Many implicit assumptions of positivist or neopositivist thought all came under attack: a misplaced reliance on formal models of scientific languages, theories, and discourse; the absence of any social or historical matrix for scientific practice; unexamined assumptions about the unity of science and its cumulative progress; unexplored oppositions between metaphysics and science; and perhaps most important, unexamined assumptions about the general superiority of scientific knowledge, particularly the physical sciences since the time of Galileo. No matter how arcane the overt controversy—say, whether Newtonian physics was a legitimate approximation of quantum physics, or whether the relativity theory was qualitatively novel with respect to classical conceptions of matter and motion—the philosophy of science debates were always questioning the primacy of scientific theories in positivist ideology, and therefore, the status and value of scientific knowledge. Kuhn's 1962 *The Structure of Scientific Revolutions* put forward an accessible, if sometimes ambiguous, array of ideas about perception, scientific communities, tradition formation, problem solving, and discontinuity in theory change supported by Kuhn's historical and scientific expertise: the actual practice of science, as revealed by its history, clearly failed to match the image that positivist philosophers and many textbook accounts claimed for it. This history was not that of obscure discoveries or marginal scientists but mainstream science at its best: the science of Nicolaus Copernicus, Galileo Galilei, Isaac Newton, Antoine Lavoisier, Niels Bohr, Hendrik Lorentz, Albert Einstein, and others.

Kuhn's model of scientific change included a temporal scheme with supporting conceptions of scientific practice. The former was roughly a two-phase description of scientific progress. In what Kuhn called normal science, relatively routine "puzzles" or "problems" are attacked using methods drawn from a largely accepted stock of experimental techniques, theoretical methods, and problem-solving approaches; these make up what Kuhn originally labeled a scientific "paradigm." A paradigm always contained contradictions and anomalies, but could nonetheless sustain progress by sticking to its fundamental vision. That particular "imperfection" is already a problem for positivist conceptions of knowledge that allow only strictly consistent theories into science. On top of that, the practices of normal science involve much craft and are not readily represented through traditional formalisms. At some point, as the resources of the paradigm become exhausted and a competing paradigm begins to emerge, and increasing significance is given to undigested anomalies, the once-normal science enters a crisis phase in which the old methods and concepts are seen as requiring substantive change.

The transition from normal to crisis phase results not solely from the intellectual merits or demerits of solutions to scientific problems but from a heterogeneous mix of intellectual content, extrascientific political or social controversies, the personal interests and abilities of participants, tacit or unconscious perceptual biases built into theoretical vocabularies, and sui generis criteria of progress. The innovations of a new approach, theory, program, paradigm, or disciplinary matrix—call it what you will, the key is that violating almost any methodological rule is possible in disposing of an old normal science, just as a normal science limits opportunities for novel criticism and innovation—give rise to a Kuhnian revolution, perhaps involving radical changes to participants as well as theories. The traditional objectification of ideas in theories results from the broadest possible conception of scientific practice. Different scientists may find themselves working in so-called incommensurable theories between which there is no straightforward translation or logical deductive relations. Instead, the scientists continue an ad hoc series of language games in formation and transition. This inventiveness generates a new paradigm even as it creates obstacles to the communication patterns typical of normal science, and thus makes crucial individuals' ability to participate in a revolutionary-like change in thinking and scientific practice. Continuing the revolution analogy, according to Kuhn, the new paradigm may make use of rhetorical and propagandistic means for ensuring the social acceptance of its ideas: through the biased rewriting of textbooks, changes in scientific curricula, and Whiggish rewrites of history.

Kuhn's ideas went through some changes over the years, especially his withdrawal of the many-faced "paradigm." The above account also does not reflect Kuhn's detailed use of history as data to support a radical theory of science's past and Kuhn's new image of it. As his book famously begins: "History, if viewed as a repository for more than anecdote or chronology, could produce a decisive transformation in the image of science by which we are now possessed." The aim, Kuhn said of his essay, "is a sketch of the quite different concept of science that can emerge from the historical record of the research activity itself" (*SSR* 1). For all their differences, Kuhn, Lakatos, and Feyerabend were all dedicated to the rich detail of science's past as the new standard for what was philosophically real and salient. The negative goal of Kuhn's message was the disposal of positivism, and supported by Lakatos and Feyerabend, he *did* decisively change conceptions of science throughout the social sciences, humanities, and science itself. "Paradigm" is even now a term of art in business culture, in spite of having run aground in philosophical waters decades ago.

The positive goals for history were less clearly achieved. Since Kuhn integrated so many ideas, it was easy for debate to turn in any direction, and without minding the primary introduction of history as the backdrop for the entire discussion. In the Anglo-American world, the direction of debate was often toward ahistorical language: of meaning invariance, reference of scientific terms, and the coherence of the idea of a paradigm's "conceptual schemes" as they were sometimes called.[1] These linguistic and analytic approaches to Kuhn mostly missed his historical perspective, or chose to see it as being grounded in an incoherent theory of language. Feyerabend, too, had limited interest in making incommensurability a "technical" philosophical concept: for him, it names a phenomenon in history that could be explored through a historical anthropology informed by a socially critical sensibility. It was ultimately less important to defend incommensurability than to recognize and understand, for instance, the great differences between the Homeric and classical Greek worldviews, or Aristotelian and modern science, and then to see how these historical interpretations also functioned as powerful counterexamples for many received notions of scientific change.

Feyerabend also saw the central issue to be the new role for history in philosophy. As he began his manifesto *Against Method,* second to Kuhn's *Structure of Scientific Revolutions* as the most important book in the debates: "History generally, and the history of revolutions in particular, is always richer in content, more varied, more many-sided, more lively and subtle than even the best historian and the best methodologist can imagine."[2] Moreover, *Against Method* is not a book lacking in theory. Feyerabend uses theories of observation in science—especially in his account of Galileo and his telescope—theories of the role of language in concept formation, and theories about major historical changes such as that between the pre-Socratics and later Greeks. But unlike Kuhn, with whom Feyerabend agrees in many ways, Feyerabend has little by way of a positive model of scientific change. Feyerabend's tools of transformation lead to a Wittgensteinian city of oddly curved streets and rebuilt buildings with all kinds of stories needed to explain how the village became a town and then city; his philosophy owes more to Homeric parataxis than to Platonic logos. The book's motto, "anything goes," means that *if* a single set of rules or governing methodology for science is desired, *then* the only rule that works is "anything goes." That is elementary, and was widely misunderstood, but what is startling is the array of historical counterexamples Feyerabend brings to bear against standard methodological advice, such as "don't allow theories in contradiction with observations," "only allow theories that can be potentially falsified," "generalize inductively from facts," "maintain a clear distinction between theories and facts," "don't give metaphysical ideas a central role in scientific theorizing," or "how one discovers a theory should not make a difference in how it is justified."

Not that these nostrums are always wrong, but Feyerabend shows that *always* enforcing such rules would eliminate much great science: these can be normatively undesirable rules. One of Feyerabend's tactics is to let history falsify methodological rules as universal criteria, but without substituting a positive doctrine like Kuhn's as a replacement. In this way, *Against Method* is a tract of Pyrrhonian skepticism, with the title resonating with the many texts of Sextus Empiricus, like "Against the Logicians," "Against the Physicists," or "Against the Mathematicians." Feyerabend is not an "irrationalist"; he is just a thoroughgoing skeptic about the foundations of knowledge who, like Kuhn, applied a varied set of conceptual tools to a shared body of historical materials. No "irrationalism," then, but another historical challenge to received thinking about scientific method, accompanied in Feyerabend's case by a Brechtian populist disrespect for authority.[3]

Again, what mattered in the end was to have made history the ultimate backdrop for any serious conception of scientific method. "All sciences are *Geisteswissenschaften*," or human sciences, writes Feyerabend in *Problems of Empiricism*; "all traditions are historical traditions"; and "the history of ideas is an essential part of scientific method."[4] Feyerabend also provided criticisms of his own against Kuhn, including that the normal versus revolutionary science distinction was hard to justify: if one looks hard enough, Feyerabend argued, not much science has the stability Kuhn attributed to normal science.[5] But in spite of such criticisms—and very different rhetorical, professional, and personal styles—there is great affinity between Kuhn and Feyerabend's attitudes toward science. Both use history to give prominent roles to rhetoric, politics, and intellectual craft; both see opportunities for unformalizable novel change; both share similar antipositivist views on the distortions created about science's past by philosophy; and in the 1960s and 1970s, both believed Western culture needed a substantive revision of its views of science.[6]

It is with respect to the implications of a new "image of science" that Kuhn and Feyerabend are best differentiated, and lead into the orbit of Popper and Lakatos. While history is paramount for Kuhn—and if his catholic and American-pragmatic adoption of a host of supporting ideas is the reason both for his success in general, and failure with specialized philosophers and antitheoretical historians—then the part of Kuhn's thinking that did him in intellectually with Feyerabend and Lakatos is his ambiguity about the normative role of his theory of scientific change. As far as science of the past is reasonably accounted for by Kuhn's descriptions, is that a good thing? Is normal science, for example, a necessary evil for progress, and is its anticritical norm actually to be endorsed in some way? To what extent is the rough schema, normal science plus problem solving → anomalies and crisis → scientific revolution → new normal science, a value-laden methodology, especially with respect to the "need" for the nuts-and-bolts normal scientist, or the technical and uncritical "problem solver" required for progress? Kuhn's inability to take a clear critical and normative stance on his descriptive categories drove Popper, Lakatos, and Feyerabend to distraction, and created a partial intellectual alliance against Kuhn. Note that the view of Feyerabend as an "anti-science philosopher," as the *New York Times* headlined his 1994 obituary, is just false; he had an obvious great love and admiration for classical science. Nonetheless, the work of Bohr, Einstein, Ludwig Boltzmann, Ernst Mach, and Galileo differs greatly, for Feyerabend, from ideological claims for science's privileged status made by philosophers or scientists themselves. Feyerabend was completely allied in his criticism of normal science and its dangers with Popper and Lakatos, and it is fair to say that Kuhn never clearly identified just what should be *done* with the new image of science. Indeed, Feyerabend provided something of a critical doctrine of science's role in society, arguing for a kind of libertarianism of ideas, adapting John Stuart Mill's ideas on freedom from *On Liberty* to the proliferation of theories in science and knowledge generally. While this political slant is not elaborated at length by Feyerabend, at least he expressly articulated his views as being about the role of science today.[7]

It may have been that Kuhn took normative philosophy of science to be another piece of positivism not worth retaining. In any case, he did not give his history an explicit normative direction, even as that applied to history writing itself. In his chapter "The Invisibility of Revolutions," Kuhn takes up one of the most controversial issues of *The Structure of Scientific Revolutions*—that of the need for nearly religious commitment on the part of participants in creating substantive progress: "The analysis now required will begin to indicate one of the aspects of scientific work that most clearly distinguishes it from every other creative pursuit except perhaps theology" (*SSR* 136). "As the source of authority," Kuhn writes of the doctrinaire commitment to a normal science, "I have in mind principally textbooks of science" (ibid.). To create effective engineers of research practice, textbooks describe and inculcate the methods to adopt and those to exclude within a rewritten history of the subject making the new practices transparent. "In the case of textbooks, at least, there are even good reasons why, in these matters, they should be systematically misleading," notes Kuhn (*SSR* 137). Textbooks "inevitably disguise not only the role but the very existence of the revolutions that produced them. . . . Textbooks thus begin by truncating the scientists' sense of his disciplines' history and then proceed to supply a substitute for what they have eliminated" (ibid.). Just as history is the great source of scientific change for Kuhn, the rewriting of history becomes its practical implementation, but Kuhn sees that more as a fact of life than an epistemological-moral problem that he now must shepherd forth.

To be fair, Kuhn recognizes the danger in what he is saying, but plays both sides of the propaganda function for history and education. On the one hand, "The deprecation of historical fact is deeply, and probably functionally, ingrained in the ideology of the scientific profession, the same profession that places the highest of all values upon factual details of other sorts" (*SSR* 138). Kuhn quotes Alfred North Whitehead's "A science that hesitates to forget its founders is lost," but that means, unfortunately, "a persistent tendency to make the history of science look [counterfactually] linear or cumulative, a tendency that even affects scientists looking back at their own research" (ibid.). As part of this process,

> when it repudiates a past paradigm, a scientific community simultaneously renounces, as a fit subject for professional scrutiny, most of the books and articles in which that paradigm had been embodied. Scientific education makes use of no equivalent for the art museum or the library of classics, and the result is a sometimes drastic distortion in the scien-

> tist's perception of his disciplines' past. More than the practitioners of other creative fields, he comes to see it as leading in a straight line to the discipline's present vantage. In short, he comes to see it as progress. No alternative is available to him while he remains in the field. . . . Inevitably those remarks will suggest that the member of a mature scientific community is, like the typical character of Orwell's *1984,* the victim of a history rewritten by the powers that be. Furthermore, that suggestion is not altogether inappropriate. There are losses as well as gains in scientific revolutions, and scientists tend to be peculiarly blind to the former. On the other hand, no explanation of progress through revolutions may stop at this point. To do so would be to imply that in the sciences might makes right, a formulation which would again not be entirely wrong if it did not suppress the nature of the process and of the authority by which the choice between paradigms is made. (*SSR* 167)

That is as ambiguous a statement as could be contrived for the normative role for history. Kuhn thought he was being descriptive and recognized difficulties with his historiography, but had no philosophical recourse.[8] With all due respect, Kuhn's historicism lacked a moral vision for the role played by history in the revised image of science. He surely did not provide anything like what might be meant by Bildung. Popper, for his part, was revolted by what he read in Kuhn, especially what he took to be the model of the anticritical normal scientist, not to mention passages like the one quoted above.[9] But as Kuhn saw, Popper really was ahistorical in practice, and so Popper's criticism of Kuhn's ambiguous normative stance was lost amid the counterflank on Popper's ahistoricism by Kuhn, Feyerabend, and Lakatos.

The discussion in chapter 5 of Lakatos's criticism of historical falsification in the history of mathematics and the irrationalism of Euclidean-deductivist style in mathematics textbooks suffices to suggest how Kuhn's Orwellian allusions would be anathema to Lakatos. But like Kuhn, and unlike Feyerabend, Lakatos would have a specific theory of the history of science: his methodology of scientific research programmes. The changed attitude toward history itself was Lakatos's philosophical center. In fact, that center is not even history; it is historiography. There is a tendency to view the philosophy of science as prescribing what should be done with science *now,* and that is partly correct. But what Kuhn, Feyerabend, and Lakatos all did was to evaluate science *past* by creating new histories combined with philosophical criticism. This activity is Lakatos's focus. In view of the sophisticated historiography already noted in *Proofs,* that should be no surprise. Yet here, that vision is conjoined not with a philosophy of learning and education but criticism, and the role for history in promulgating "images" of science. Given, for example, Kuhn's conception of textbooks and their propaganda function, what should the role of history be in teaching about science's current state? Are the categories of normal science and crisis useful for the goals to be achieved? Kuhn also criticized strict Popperian falsification as historically inaccurate, but what then are the roles for contradictions in theory change? Is there a middle ground between the formality of logical positivism and a nonspecific psychology of discovery? What should be said of specific historical achievements—such as the emergence of the relativity theory or the wave theory of light, or major experiments like Michelson and Morley's on the speed of light—and how should their received accounts be evaluated? Just as we judge social and political history, so too we can make evaluative critical judgments, not just passive endorsements, of science or its history. One may disagree with Lakatos's answers, but at least his philosophy of science faces such historical-cum-historiographical questions head on. An algorithmic comparison of theories it is not.

Lakatos ends up creating an account both of how history is built into science itself and how methodology grows out of the history of science. "Methodology is wedded to history," Lakatos says, "since methodology is nothing but a rational reconstruction of history, of the growth of knowledge" (*MSE* 178n3). In short, Lakatos provides a dialectically developed conception of scientific change, oriented around criticism and historiography. Like *Proofs*, Lakatos demonstrates all his ideas in medias res, thereby carrying out several projects at once: an immanent criticism of Popper, the creation of his own methodological theory, the writing of several histories, and the development of a historiographical theory addressing all that. His primary goal is the development of a historical account of scientific method through a critical elaboration of Popper's philosophy. History is not assumed as another dogma but developed from Popper's ahistorical philosophy as the empirical resource by which methodologies can be evaluated. As an end, that agrees with Kuhn's use of history to refute Popper and positivist representations of scientific theories. But just what goes wrong with Popper? And how will a historicist philosopher of science finally position *himself* in the critical stream of time? Contrasting himself in *Criticism and the Growth of Knowledge* to Popper and his other critics, Kuhn argues that

> the explanation [of scientific progress] must, in the final analysis, be psychological or sociological. It must, that is, be a description of a

> value system, an ideology, together with an analysis of the institutions through which that system is transmitted and enforced. Knowing what scientists value, we may hope to understand what problems they will undertake and what choices they will make in particular circumstances of conflict. I doubt that there is another sort of answer to be found.[10]

True enough, there was nothing in Anglo-American or continental philosophy of the 1960s or 1970s that would offer "another sort of answer." It is just such an alternative, though, that Lakatos wants—an approach to philosophical history of science based on a historically variable conception of human reason, or at least scientific criticism. The activity on which this philosophy is centered is the creation of the histories Kuhn and others were busy writing.

The central Hegelian aspects of theory development and choice on which Lakatos focuses are dealing with contradictions, much like Kuhnian anomalies, and the backward-looking, hindsight role for historical reinterpretation within science. As we saw in *Proofs* earlier, Lakatos implicitly uses the idea of self-formative Hegelian Bildung to describe the mathematical present as history and explain a mitigated-skeptical notion of mathematical progress. Here, that self-formative logic is expressed through an account of how histories of science are written and become part of our conception of science. And as Lakatos builds his historicized understanding of scientific change out of Popper's philosophy, in addition to showing how to get from Popper into history, he supplies an active demonstration of the very methodology he creates as "science." The criticism of Popper is dovetailed with Lakatos's normative histories to create a historiographical theory, all of which constitutes, in my view, Lakatos's changing logic of scientific discovery, meaning a methodology that implies its own historical and transient status. As Feyerabend noted in his obituary for Lakatos, in modern science there is a constantly expanding horizon of *facts.*[11] The methodology of scientific research programmes shows that there can be a constantly expanding horizon of *method* as well.

This page intentionally left blank

7 | A Historiographical Toolkit

Lakatos's philosophy of science is essentially contained in two essays: "Falsification and the Methodology of Scientific Research Programmes" (*FM*) and "History of Science and Its Rational Reconstructions" (*HS*).[1] The two papers amount to a compact and effectively complete book. The longer "Falsification" essay leaves open questions explicitly answered in "History," as well as providing historical case studies to which the historiographical theory of "History" applies. Several cues in the first "Falsification" essay, published originally in *Criticism and the Growth of Knowledge,* also suggest that it was constructed to develop the historiographical problems explicitly discussed in the "History" sequel.[2] The two pieces form a continuous and continuously reflective organic whole, and should be read as such.

The methodology of scientific research programmes is not baldly "proposed" or simply stated by Lakatos. The methodology is developed dialectically as the synthesis and culmination of a sequence of successively more sophisticated and powerful methodologies. Lakatos's philosophy of science, his "methodology of scientific research programmes," is an elaboration of Popper and a merciless attack by way of historicizing Popper's ideas. Also included are Pierre Duhem and others' "conventionalism," and surprisingly even a residue of logical positivism. The last is the idea of a demarcation criterion itself, representing the original goal of creating normative criteria to distinguish science from its other.

Such an integrative and ostentatiously unoriginal philosophy is characteristically Hegelian, and its constructive and critical dimension is entirely missed by just restating what a research programme is. Lakatos's concept of a research programme is itself a minor history of twentieth-century method, centered on Popper, analogous to Hegel's development of his philosophical system as a critical summary of philosophy's past up to his own time, and weighted heavily toward Kant, Fichte, and Schelling.[3] Resisting "ready-minted" philosophy is even one of Hegel's key criticisms against romanticism; for Lakatos, his own philosophical writing is like a mathematical proof in that there's no substitute for the real thing. Also like Hegel, Lakatos's research programme "system" is intended to be applied to any discipline with value-laden criteria of progress, such as economics, just as Hegel applies his system to the history of politics, religion, and art. In this way, Lakatos's approach, like Hegel's, is also "totalizing" through its application to different intellectual domains.

Another example of this historical and synthetic approach is Marx's critique in *Capital* of classical political economy, which is built out of the economic categories and theories of Adam Smith, David Ricardo, and many others. Marx's critique has many sides to it, but the most relevant here is his approach of jointly historicizing economic categories and political economy itself. That Marx intended to explain the historical specificity and transience of economic categories, their dependence on distinct social relations of labor, exchange, and distribution, and certainly their nonnaturalness, is one of the primary and lasting contributions of Lukács's Hegelian interpretation of Marx's work: just as Hegel historicized philosophy, Marx historicized political economy. For Lakatos, the topic and purpose are vastly different: the goal is to historicize categories of scientific criticism and the philosophy of science. Lakatos will also end with an extremely Hegelian historiography, rather than a Marxian social theory or Hegelian philosophy of history. But the method toward that goal is much the same: an immanent critique of major received views of the era that literally becomes a new theory and a historical account of its own progress.

Much in Popper, then, is taken as a starting point. First is the goal of wanting to demarcate science from other types of knowledge, of normatively characterizing what is useful in science's success. Popper's famous answer, expressed originally in his *Logic of Scientific Discovery,* was that against the positivists, scientific knowledge could never have, nor was in need of, a foundation, neither through logic, the senses, nor cognitive faculties. Attempts at inductive logics, a role for positivist sense-data, Kantian theories of experience and consciousness justifying synthetic a priori knowledge, semantic theories intended to ground science in a foundation of meaning and reference, are all from this perspective fundamentally ill conceived. Logic, though, for Popper, even if it cannot inductively support the universal statements of science, does imply that scientific conjectures are falsifiable: they are potentially refuted through a single observation or experiment inconsistent with it. Scientific method can assume logic as given and unfalsifiable while not implying any ultimate foundation for conjectural scientific knowledge. Science progresses for Popper through the "negative learning" Lakatos took up in *Proofs:* through the proposal of falsifiable conjectures, their testing, possible disposal, and then replacement. While Popper's philosophy of error, of "learning from our mistakes," and the basic logic of conjectures and refutations was taken for granted in *Proofs,* it now is subject to historical and critical scrutiny. The historical perspective shifts from the "forward-looking" approach to learning and history of *Proofs,* to the "backward-looking" assessment of scientific theories: when they are progressing or degenerating, and whether they even count at all in Popper's demarcation game as science. This meaning for methodology is all-important for Lakatos: "The term 'normative' no longer means rules for arriving at solutions, but merely directions for the appraisal of solutions already there. Thus methodology is separated from *heuristics,* rather as value judgments are from 'ought' statements" (*HS* 103n1).[4]

Lakatos begins his critique of Popper as might an ancient skeptic, with his opponent Popper's problem of characterizing "science" and also "good science." This "demarcation game" was adopted earlier by twentieth-century positivists as part of their philosophical hygiene intended to cleanse the world of illegitimate knowledge claims.[5] The demarcation game is a modern variation on an old philosophical tradition beginning with Socratic-Platonic inquiries asking, for example, what is virtue, knowledge, or justice. Kant's more recent demarcation strategy was to limit the claims of reason, or Vernunft, against his version of cognitive-scientific understanding, or Verstand, and to limit both of these vis-à-vis cognitive experience. Lakatos's innovation is that answering the question "What is science?" may involve historical explanations of science and the criticism of histories. Rules of demarcation are turned into a historical topic and targets of criticism as implicit historiographical rules. In this way, the normative quest of positivist philosophy is continuously transformed by Lakatos into normative historiography by way of Popper.

The Series of Poppers

This goal is achieved by Lakatos's overall objective of constructing the methodology of scientific research programmes through a sequence of progressively stronger methodologies:

$$\text{Popper}_0 \rightarrow \text{Popper}_1 \rightarrow \text{Popper}_2 \rightarrow \text{methodology of scientific research programmes}$$

The subscripts are not a formalism but a shorthand Lakatos uses to disaggregate and combine aspects of the "real Popper's" ideas into a progressive pattern culminating in the theory of research programmes. The approach is similar to, but less extreme than, the caricatured shapes of mathematical history used in *Proofs.* The real Popper is mostly Popper$_1$, and parts of Popper$_2$, while Popper$_0$ is a positivist misrepresentation of Popper nonetheless useful for Lakatos's reconstruction.

Each of the Poppers includes a demarcation criterion for science and characteristic instances of what meets or fails to meet that standard of knowledge. Popper$_1$ is the closest to the familiar Popper: for a theory to be scientific, it must have a set of potential falsifiers, meaning the theory-laden observations, specified in advance, that would count as refuting the theory and leading the theory to be abandoned were they to be observed. To use Popper's famous examples of vulgar Marxism or versions of Freudian theory as *un*scientific knowledge, potentially falsifying conditions of this type are never identified and there is always a way to rationalize away the worst clash with reality, be it the failure of a centralized socialist economic plan or therapeutic regime. In this way, such knowledge claims are fundamentally dishonest. For Lakatos, too, methodologies define codes of scientific honesty, which, Lakatos will eventually show, is reflected in the more or less falsified histories to which they give rise. For Popper$_0$, a scientific theory is not one that is foundationally justified, yet it should have some potential falsifiers; the difference from Popper$_1$ is that the potential falsifiers of Popper$_0$ are not theory-laden: Popper$_0$ assumes, like the positivists, an observational foundation against which theories may be conclusively falsified.

For Popper$_0$, though scientific theories cannot be proved, they can be conclusively refuted through a completely stable observational basis of falsifying knowledge. Popper$_1$ and the real Popper, in contrast, are what Lakatos calls "methodological falsificationists" who recognize, as Popper wrote, that observations or experiments are always "interpretations in the light of theories" (*LSCD* 59n*1, 107n*3) and an associated network of "conventional" decisions assigning truth-values to theory-laden observational statements.[6] Earlier philosophers of science such as Duhem recognized the impossibility of inductive proofs, and instead considered conventionally assigned truth-values to scientific generalizations, such as Newton's laws, Maxwell's equations, or phenomenological laws of heat. Such generalizations cannot be proved using logic or specially designated observations, although as the conventionalists say, based on the accumulation of evidence and coherence, truth-values may be conventionally decided. Popper's innovation was to combine conventional truth assignments to theory-laden "potential falsifiers" with his new falsification criterion, with inductively supported generalizations being recast as fallible theories.

Thus, the positivist idea of a demarcation criterion is preserved by Lakatos, while its error of trying to justify knowledge is reversed in Popper_0, who recognizes that criticism and fallibilism drive progress; then the error made by Popper_0 of not recognizing that falsifying observations are theory-laden is resolved in Popper_1, while retaining falsification. Each of these early shapes of Popper is thereby both negated and superseded by the methodology of scientific research programmes, providing neat and quite intelligible examples of Hegel's Aufhebung, here meaning the joint rejection and retention of different aspects of predecessor methodologies.

The criticism of Popper_1 is two-pronged. One prong is abstract and is provided by Lakatos's holistic, Duhem-Quine-style ceteris paribus thesis. That means that if no theory is refuted in isolation, but only against an extensive network of background assumptions, or ceteris paribus clauses potentially open to refutation, then for the falsifications of Popper_1 to remain stable, confrontations between theory and observation require a large number of essentially permanent decisions about which pieces of knowledge to "hold fixed," and which to make subject to change. Otherwise, theories can potentially pop in and out of "refuted" status and render progress unintelligible. Defining these decisions is a serious challenge for Popper_1 and the real Popper—a challenge that ultimately cannot be met because the condition is just too strict: it is almost as if all scientific knowledge was based on the Neyman-Pearson model of comparative statistical hypothesis testing. The problem is that if the holistic arrow of refutation can in principle be redirected anywhere in the network of knowledge, progress cannot be localized to isolated confrontations of theories and theory-laden observations. That difficulty is expressed in the real Popper through a complicated hierarchy of decisions needed to demarcate a "theory under test" from the welter of ceteris paribus clauses.

It is the second prong of Lakatos's criticism of Popper_1 that begins the march toward history. Lakatos asks the question, What could count as finally falsifying Popper_1 as methodology, meaning to reject *it,* just as Popper_1 would reject a falsified theory? That is, looking at Popper_1 as yet another shape of knowledge, what internal criteria might help overcome its errors? Wouldn't it just be a theory-laden refutation of some kind, since *those* effectively define the standard Popper_1 offers as knowledge? History, real science, is what Popper_1 is trying to match—Einstein, Bohr, and Newton are "in," and Marx and Freud are "out"—and it is this implicit historical yardstick or standard against which Popper_1, as a criterion of knowledge, should be measured. Even the real Popper cites famous episodes from the history of science as representing his falsification criterion in action. In the long "Falsification" paper, Lakatos here makes his first sustained use of historical counterexamples and initiates a role for the so-called actual history of science. From history, Newtonian physics was not "given up" because of known problems between the theory and Mercury's observed path at its perihelion, now accounted for by the relativistic influence of the sun's gravitational field on Mercury's orbit. It is also unclear what conditions could have been specified in advance of relativity theory as a set of crucial experiments for rejecting Newton's laws, not to mention many other scientific theories. But a sophisticated history is still beyond reach at this stage. Even if Popper$_1$ is apparently falsified by history, there's no basis yet for a historicist perspective with the horsepower needed to write history with any finesse. This conception of history as *only* falsifying is too simplistic, yet it is here just right: it is a "naive" historicism corresponding to the "naive falsificationism" of Popper$_1$. Lakatos describes this historical yardstick four times, while also noting proleptically that he will have to proceed in assessing Popper's methodology without a theory of how to do it.[7] His sequel explains just that.

The role of Popper$_2$, then, is the positive introduction of time, historical character, and temporal reflection missing from Popper$_1$ and the real Popper into the assessment of scientific theories. That truly deep idea of post-Kantian philosophy is made all the more interesting here given the simplicity and analytic elegance of Lakatos's approach. Instead of concentrating on a single theory T confronting the world, Lakatos proposes as the basic unit of appraisal for Popper$_2$, who he labels the "sophisticated falsificationist," a series of theories $T_1 \rightarrow T_2 \rightarrow T_3$. . . . So there are in history successive versions of, for example, the wave theory of light: not just Thomas Young's theory, but before that, Newton and Christian Huyghens's wave theories, and after Young's, Augustin Fresnel's wave theory; or the succession of Lorentz's ether theories following James Clerk Maxwell's electrodynamic laws; or the succession of phlogiston theories, and parallel "dephlogisticated air" and oxygen theories; or the series of molecular and phenomenological theories of heat; and so on.

What is important now are successive changes between theories, not just their falsification, and separating the rejection of a single theory from the overall appraisal of a sequence of theories. First, Popper$_2$ might reject a theory not because it is falsified but because an improved theory supersedes it. But you may also hang on to a theory, even if it has many known anomalies or contradictions, for want of a better one: for Popper$_2$, there is no rejection of a theory without an improved theory to take its place. "All theories are false," meaning that anomalies or contradictions are ubiquitous, implies that the relevant question is not "Do I now accept or reject *T* as true?" but rather, "Is *T*′ an improvement over *T*?" Changes between theories are depicted by Lakatos as "problemshifts," and the desirable problemshifts are ones in which old observations are explained, some novel facts are predicted that are unanticipated by predecessor theories, and some of these novel predictions are corroborated or confirmed. A problemshift is theoretically progressive if the new theory predicts new theory-laden facts, and empirically progressive if some of these predictions are also corroborated. For example, let *T* be Newton's physics, *T*′ Einstein's general theory of relativity, and the observation be the curvature of light paths around large masses such as the sun. Then,

> Einstein's theory is not better than Newton's because Newton's theory was "refuted" but Einstein's was not: there are many known "anomalies" to Einsteinian theory. Einstein's theory is better than—that is, represents progress compared with—Newton's theory anno 1916 (that is, Newton's laws of dynamics, law of gravitation, the known set of initial conditions; "minus" the list of known anomalies such as Mercury's perihelion) because it explained everything that Newton's theory had successfully explained, and it explained also *to some* extent some known anomalies and, in addition, forbade events like transmission of light along straight lines near large masses about which Newton's theory had said nothing but which had been permitted by other well-corroborated scientific theories of the day; moreover, *at least some* of the unexpected excess Einsteinian content was in fact *corroborated* (for instance, by the eclipse experiments). (*FM* 39)

By Popperian$_2$ standards, a contradiction is neither necessary nor sufficient for eliminating or falsifying a theory. Not necessary because a progressive problemshift need not contain any Popperian$_1$ refutations. Not sufficient because a theory is not regarded as falsified until a new one can replace it; by this alone, the difficulties in "designing" a crucial experiment to segregate ceteris paribus conditions from the theory at test disappear. An attempt to holistically replace *any* part of science can be made, but it is not progressive if it doesn't improve on the past. Popper$_2$ also begins to change the modern conception of the history of science. For both Popper$_2$ and Lakatos, it now is a category mistake to apply the term "scientific" to a single theory, and for Lakatos, it was one of Popper's basic confusions to confound "theories" and "series of theories." Indeed, writes Lakatos, Popper's "ambiguous usage led to such confusing formulations as 'Marxism [as the core of a series of theories or of a research programme] is irrefutable' and, at the same time, 'Marxism [as a particular conjunction of this core and some specified auxiliary hypotheses, initial conditions and a ceteris paribus clause] has been refuted' " (*FM* 34n5).[8] Popper$_2$ separates appraisal from rejection: theories that are part of degenerating problemshifts can be appraised, but not yet rejected, and still be within "science," just not "progressive" science. As shown in chapter 11, this is just the case for Marxian economics.

Falsification now is a multiple relation between the theory-laden empirical basis from Popper$_1$, the competing theories that interpret and explain these observations, and the kinds of growth between successive problemshifts: "Falsification can thus be said to have a '*historical character.*' . . . 'Crucial counterevidence'—or 'crucial experiments'—can be recognized as such among the scores of anomalies only *with hindsight,* in the light of some superseding theory" (*FM* 35, 36). The "sophisticated" falsification criterion of Popper$_2$ is an austere Hegelian criticism: the criterion is explicitly retrospective and the new T' may still abound with anomalies—it may be false—but T' is progressive if it explains its predecessor's positive content and something more. By simultaneously raising the importance of actual history to, as it were, refute Popper$_1$, diminishing the force of contradictions and introducing a spartan historicism, Popper$_2$—mostly Lakatos's own creation[9]—provides a flexible, protohistoricist model of scientific change. Popper$_2$ also implicitly implies that Popper$_1$'s resistance to historical thinking is due to an overreliance on classical logic, the law of noncontradiction, and a simplified view of the working role of contradictions in scientific practice.

Popper$_2$ is still just a way station toward the methodology of scientific research programmes. The central problem with Popper$_2$ is characterizing the coherence in a series of theories and their rejection. What is constant through the changes brought about by problemshifts? How, for example, are the various wave theories of light, or phlogiston theories, or atomic theories of heat, recognized? Where does a new theory in a series come from?—that is, is that discovery intrinsic to the series or serendipitous? What prevents tacking together theories, meaning adding on an unrelated, low-level hypothesis to a theory, which by Popper$_2$ could constitute progress? Finally, what does it mean to reject a whole series of theories, rather than a single one? Lakatos's answer to these questions is provided by a position that, unsurprisingly, is continuous with the sophisticated falsificationism of Popper$_2$. Popper$_2$ is completely absorbed as the methodology characterizing the local criticism that coheres as a research programme. Thus, the transition to the methodology of scientific research programmes is partly what Popper$_2$ is about: pure positive progress, an improvement without a refutation.

The series of theories making up a research programme is unified by the programme's "hard core," providing the fundamental principles on which the programme is based. The hard core may change over time, but at a certain point, giving up the core or substantively changing it means to have rejected or tabled the programme. Examples of hard cores include Newton's three laws of motion plus the theory of gravitation for classical mechanics, or the theory of chromosomes plus the information-processing role of DNA as basic genetic material, or Maxwell's equations for electromagnetics. Most theories can readily be thought of as research programmes: the wave theory of light, the corpuscular theory of light, nineteenth-century ether theories, electromagnetic theory, relativity theory, plate tectonics in geology, early quantum theory, the Eldredge-Gould theory of punctuated equilibria in evolutionary theory, theories of global climate change, various economic theories, even the development of fields of mathematical research.[10] Research programmes offer a different view of familiar materials via the toolkit of research programme components used to write their histories. Theories that implement and elaborate the hard core are part of the programme, so the same person can work on multiple, competing, or inconsistent programmes. Many "Newtonians," for instance, proposed changes to the gravitational law in order to accommodate anomalies in the orbit of Uranus, later explained by the Adams-Leverrier discovery of Neptune.[11] The hard core is not tested or even used directly, and need not be falsifiable. Rather, it is implemented via the programme's "positive heuristic," which suggests methods or a plan for articulating and revising the programme forward. The positive heuristic helps generate a series of testable theories and models that then form the programme's "protective belt," which in turn is aided by supporting mathematical theories, instrumentation and observational theories, and auxiliary hypotheses from allied fields. Given the significance of temporal change in a research programme, the positive heuristic can be more important than the hard core.

Theories in the protective belt may be compared according to types of problemshifts between them, as contained in Popper_2. If the theories used to generate a problemshift are largely imported from outside the programme without a coherent tie to its positive heuristic, or if the theory is softening the assumptions of the hard core, then the problemshift may be "ad hoc," though that could be remedied by a later change.[12] Theories also are used to explain, reinterpret, or defuse anomalies or contradictions that threaten them, sometimes just to fight off competitors, but possibly to provide additional, if not novel, corroboration as well. The theories of the protective belt may be refuted closer to the style of Popperian falsification, though they still may harbor considerable imperfections. As for Kuhn, anomalies or outright contradictions may be ubiquitous for a research programme. Progress can occur *without* falsification if a more powerful theory happens to be available. Confrontations between theory and observation also occur only via theory-laden observations or facts. Just as the hard core may evolve, so too may the positive heuristic change, as new techniques evolve for elaborating the programme, such as methods for modeling small particle motions, planets, or fluids, all within classical mechanics as a whole. So too, the modern theory of evolution can be applied on different timescales, or at different levels of biological organization and complexity, giving rise to theories and heuristics for subjects ranging from the interpretation of the Paleolithic record to animal husbandry.

Hence you have science, consisting of research programmes; and good and bad science, consisting of progressive and degenerating research programmes. Typically, it is only from a long view of cumulative development, possibly including critical phases of both progress and degeneration, that a programme can be adequately assessed, and even then only in comparison to its competition. To take an extreme case, the atomic theories taken for granted today existed in some form since Democritus, but competed weakly against continuous theories of matter until almost the beginning of the twentieth century. Scientific progress for Lakatos is long-run, theory-laden, and retrospective research programme progress. Nothing ever conclusively dictates rejecting a research programme, or how to ensure programme progress. As a normative tool, it is always backward-looking, and even then at best suggests the relative risks of continuing or tabling a research programme at a given point in time.

That, in a nutshell, is the logic of research programme categories. The cash value of the methodology is in the histories it generates, just as much of the value of *Proofs* is Lakatos's account of nineteenth-century mathematical methods. Lakatos includes several relatively short case studies in his original essay, "Falsification"; indeed, his critique of Popper smoothly segues into the writing of history using research programme categories, and then to a historiographical theory. Chapter 9 will look at Lakatos's two most important examples: his account of Bohr's early quantum theory and the Michelson-Morley experiments on the speed of light. Both help explain science past while making strong philosophical points. The lessons are completely anti-Popperian. In the case of Bohr, the lesson is the ability of science to make progress on thoroughly inconsistent foundations, while for the Michelson-Morley experiments the issue is the inability to specify, as Popper wants, conditions of refutation in advance, and consequently, the crucial role of historical hindsight within scientific reason. Both of these ideas notably appear in *The Structure of Scientific Revolutions,* but they are now presented as central to a quite different conception of historicized science.[13]

More representative of the kind of histories that Lakatos's historiographical programme is intended to generate are ones written by Lakatos's students and published shortly after his death.[14] These studies truly continue Lakatos's work from philosophy into history of science proper, which as we will see, is exactly where his criticism of Popper takes him. Regardless of *what* is said in these histories, their emergence out of a philosophical debate is testimony to Lakatos's powerful historical vision. These applications are, dare we say, examples of normal historiography using research programme concepts.[15] These histories too, including Lakatos's, sometimes contain considerable imperfections of their own, but they are also novel correctives to much that came before them, and demonstrate like nothing else in Anglo-American philosophy the practical execution of a philosophical-historical research programme.[16]

Thomas Young and the Wave Theory of Light

The English prodigy Thomas Young, who helped decode the Rosetta stone among other achievements, is known for reintroducing the wave theory of light through a series of demonstrations conducted around 1800 on interference phenomena. One experiment involved directing a point source of light first through a narrow slit, then through a second set of double slits behind which an interference pattern of alternating light and dark bands can be seen on an observation screen. This first experiment was considered by Young, and frequently in history, to demonstrate the wave nature of light by showing that cancellation effects can occur when light is "added" to light, with wave troughs canceling wave peaks. A second experiment, considered often to be crucial against corpuscular theories of light, involved inspection of the fringes inside and outside the shadow cast by a narrow object in front of a light source; when the light was allowed only to be cast on one side of the object, the fringes disappeared, demonstrating, it was said by Young and others, that the original fringes were caused by interference by the light waves traveling around the two sides of the object.

These two experiments are among the most famous in the theory of light, but it has also long been observed that they failed to effect a switch from the then dominant Newtonian corpuscular theory of light. It was only after the 1820s and continued work on the wave theory by Augustin Fresnel in France that the wave theory was accepted. Did the decades-long delay mean that the scientific community, or subgroups in England or France, unreasonably resisted the introduction of a new and more powerful theory? Such external explanations have appeared often in descriptions of Young's initial failure to convince his contemporaries that the wave theory was shown to be true through his demonstrations. One account is that the corpuscular theory, having been substantially developed by Newton, was supported by the authority and chauvinistic entrenchment of Newtonian methods.[17]

In his essay in *Method and Appraisal in the Physical Sciences,* John Worrall identifies two research programmes in the physical theory of light around 1800: a wave programme and a corpuscular one. The hard core of the wave optics programme was that "light is a disturbance in an all-pervading elastic medium" (*MA* 136), later to be called the "ether," and whose positive heuristic was to explain optical phenomena through the mechanics of such an elastic media without invoking any unknown forces. This programme was not Young's invention but considered a long-standing and well-known approach. Several corpuscular theories, in contrast, were based on the idea that light is made up of particles or corpuscles emitted from luminous objects, and that these corpuscles obey the laws of Newtonian mechanics. The hard core and heuristic of the corpuscular programme in this way complements those of the wave programme. Rectilinear motion on the corpuscular account followed immediately from Newton's first law of motion, since any deviations from a straight-line path would have to be attributable to some external force. Another corpuscular-style conclusion was that observed different degrees of refraction for different colors was due to differences among the corpuscles themselves, since mechanical laws imply only one resultant force was acting uniformly on the surface of the refracting body.

Neither the wave nor corpuscular programme was developed with great quantitative sophistication, so their comparison depends very much on which qualitative explanations could be provided, and with what relative strength. The corpuscular programme had an inherent advantage in its ability to utilize existing mechanical theories, with the wave programme depending on heuristics drawn from the much less developed mechanics of elastic media. The best-known results from the corpuscular theory were given by Newton in his *Principia,* including predictions that light paths reflecting off a plane followed an incidence-reflection sine law, analogous to that for small particles bouncing off a surface. Newton was also, for Worrall, the greatest contributor to the wave programme, even compared to Newton's contemporary Christian Huyghens, who founded the modern wave theory of light. Among Newton's contributions, later developed fully by Fresnel, was the idea that light waves consisted of regular periodic undulations instead of irregular pulses, as suggested by Huyghens. Remarkably, Newton also created the corpuscular programme out of dissatisfaction with his own wave theory. The difference in their strength to Young's time may be attributed to the comparative power of the two programmes' heuristics, with the more complex mechanics of waves still needing the boost they would get later from French mathematicians such as Fourier and Poisson.

How well did Young use waves to explain light? For rectilinear motion, Young attempted a wave-theoretic explanation, but did not succeed, as Huyghens and Newton had failed as well. Young effectively treated light as rays except when they crossed, and then used wave properties to explain their interaction (*MA* 144). Divergences from straight paths were used to account for the passage of light through an aperture, such as a slit, but did not entail what should happen in the two-slit experiment. In this case, Young had to carry out an actual experiment to decide what had occurred, rather than deducing the result from a theory. Similarly, his diffraction experiment must actually be observed, and then the effect explained backward in terms of interference of light sources, rather than predicting the observed diffraction pattern from the experimental setup; Worrall also compiles evidence that Young never carried out the two-slit experiment at all (*MA* 153). All this, then, serves to make Young's work, Worrall argues, characteristic of a degenerating contribution to a research programme. The qualitative features of Young's two famous experiments were already known; they were not predicted by Young's original theory and his subsequent theories mostly explain them in ad hoc ways.

That is a devastating revisionist judgment—namely, that Young's contribution to the wave programme was almost nil. By the 1820s, Young had developed a reputation as founder of the new wave theory, with Fresnel believing that his own work owed little to Young, and Worrall's research programme description shows Fresnel to be largely justified.[18] The scientist cited often for his "crucial" two-slit experiment apparently made a modest or negligible contribution to the revival of the wave programme begun and abandoned by Newton.[19]

Phlogiston and Oxygen

The history of science includes many theories and concepts quite different from ones that are used or believed to be true today. One such theory is that based on the "imponderable" substance phlogiston. Phlogiston was named by the German chemist Georg Stahl, and from about 1750 to the end of the eighteenth century, many chemists considered phlogiston to be the "inflammable principle" released from materials on burning or in combustion processes generally. For example, since many metals turn to ash when heated, the metals were hypothesized to contain phlogiston transferred to the ash. The smelting of ores using charcoal resulted in metals, so the charcoal was conjectured to be a means of supplying phlogiston to the ores.

These mistaken phlogiston explanations can be readily translated today into basic chemistry; such is the benefit of hindsight. For much of the latter eighteenth century, phlogiston theory provided the best theoretical means for explaining numerous chemical phenomena, at least according to Alan Musgrave's research programme study. In qualitative terms, phlogiston theory was reasonably comprehensive and used without glaring inconsistencies up to the time of the great French chemist Antoine Lavoisier. Several phlogiston theories were developed by Joseph Priestly and Henry Cavendish—addressing whether phlogiston was a so-called imponderable substance that could not be isolated and measured, or whether phlogiston was "inflammable air," or hydrogen.

Lavoisier is famous for deposing the phlogiston theory through a series of experiments involving oxygen conducted in the 1770s and 1780s. These experiments are often considered to have decisively refuted the phlogiston theory in classic Popperian style by demonstrating the existence of oxygen being consumed through combustion, rather than phlogiston being released from a burned substance. Phlogiston is such an archetypal *rejected* idea it seems that one hardly needs a sophisticated methodological portrait of its demise and oxygen's ascent. But no single experiment of Lavoisier's was definitively negative against phlogiston. Phlogiston proponents managed to provide some confirmed predictions and counterexplanations to Lavoisier's, which themselves evolved through several incorrect versions accompanied by their own anomalies, and the phlogiston theorists also supplied much of the novel experimental basis that Lavoisier would reinterpret using oxygen theories.

The hard core of the phlogiston programme consists simply of taking phlogiston as the inflammable principle contained in burning substances, with the heuristic of accounting for various experimental results using the qualitative "combinatorics" of phlogisticated and dephlogisticated air, fixed air or carbon dioxide, inflammable air, nitrous air, and so on. Lavoisier's programme was based on the complementary premise that the inflammable principle was instead to be found in the ambient air. Positive qualitative predictions were created using phlogiston. Henry Cavendish proposed that when a metal like zinc or tin was immersed in an acid, a highly "inflammable air," or hydrogen, was released unchanged into the air. Cavendish argued that if the phlogiston had been released, then the oxides or ashy substances formed when metals are burned, or calxes, when immersed in acids would form salts, but not release the inflammable air; and this did occur. This was the experiment leading Cavendish and then Priestly to identify this inflammable air with phlogiston.

There were problems with this particular identification, yet it gave chemists the confidence that phlogiston was, as put by Richard Kirwan in his 1789 *Essay on Phlogiston,* no "mere hypothetical substance, since it could be exhibited in an aerial form in as great a degree of purity as any other air."[20] In 1772, Lavoisier famously conjectured that it is "a prodigious quantity of air that is fixed during the combustion and combines with the vapors," but he initially thought the relevant component was Joseph Black's "fixed air," or carbon dioxide.[21] So Lavoisier in 1772 could not be said simply to get it *right,* that oxygen is being consumed in combustion. In 1774, Pierre Bayen reported that burning mercury calx would yield Black's fixed air, which would have refuted Lavoisier's current theory. Priestly, at the time, used a theory in which airs ranged in the amounts of phlogiston they contained, and in repeating and correcting Bayen's work, found that the air produced was, as Priestly described it, "dephlogisticated air," being so pure as to take up phlogiston more readily than other airs. Priestly visited Lavoisier in Paris, telling him of this new discovery. Lavoisier repeated the experiment, and after a brief exchange with Priestly, went on to publish his explanation of it.

Yet dephlogisticated air was still not yet proposed as the basis for a new theory. Priestly read Lavoisier's publication and pointed out that Lavoisier's gas was more than common air; it was dephlogisticated air, as could easily be shown by tests, such as timing how long a mouse survived in a closed jar with only the gas to breathe. Lavoisier took a year or two to change his approach, with the explicit claim that the "purest part of air," or oxygen, supports combustion appearing only in 1778. Lavoisier, Musgrave notes, predicted oxygen without discovering it, while Priestly, who had discovered oxygen, taught Lavoisier how to discover and explain it. The event, the "discovery of oxygen," was created at the intersection of two competing chemical research programmes, has no single associated "discoverer," and is the combined accomplishment of two programmes' competing theories, experiments, and reinterpretations.[22]

Similarly, during the early 1780s, Cavendish explained how a dew that formed on the inside of combustion vessels when inflammable airs were burned might be the residue left after phlogiston is released. Cavendish found that the dephlogisticated air, or oxygen, and inflammable air, or hydrogen, disappear in the formation of the dew in a ratio of about 1 : 2. Dephlogisticated air, Cavendish conjectured, was water deprived of phlogiston. Cavendish's work was communicated to Lavoisier, who, as he had done before with Priestly, reconducted and reexplained the experiment. Lavoisier saw that water is not an element but composed of inflammable as well as dephlogisticated air, and borrowed or stole Cavendish's observation of the 2 : 1 proportion in weights. Analogous to what occurred with Priestly, Cavendish in the phlogiston programme had synthesized water and Lavoisier in the oxygen programme explained it.

Through the 1770s and into the 1780s, the phlogiston programme was still active, being used to explain that acids were formed when certain substances were burned and deprived of phlogiston, while Lavoisier would contend rather that acids were formed by the substances combining with oxygen. Again, Lavoisier borrowed experiments from Priestly yet reinterpreted the outcomes, saying that "the purest part of air" always entered into the formation of acids, and coined the term "oxygen," meaning "acid generator." This new theory predicted acid formation from the oxidation of organic compounds, and led to the discovery of dozens of new acids when only a handful had been known before. Oxygen explained acidic properties, Lavoisier thought, but he could not extract oxygen from muriatic, or hydrochloric acid, which contains no oxygen, and attributed the failure to his experimental technique. In 1808, after phlogiston theory had been given up by all, Humphrey Davy would show Lavoisier to be wrong about muriatic acid and chlorine, and Davy lightly considered the idea of resurrecting phlogiston.[23] Ironically, Lavoisier's great discovery is misnamed: "pyrogen" or "fire generator" might have been more appropriate.

By the mid-1780s, the phlogiston programme was recognized by almost all as one that continued to formulate post hoc explanations to accommodate known phenomena, in contrast to Lavoisier's consistent success. Did the phlogiston programme drag on too long? It *could* have been given up earlier, and in the early 1780s, Priestly argued some English chemists back from support of the French approach. But at just what point rejection was definitively justifiable is hard to say. Lavoisier's great 1789 text, the *Traité élémentaire de chimie,* synthesized numerous experiments and created a distinctively modern chemical perspective, including the use of systematic chemical nomenclature. In spite of its usefulness, however, the work did not provide fundamentally new factual content.[24] The suspicion that the English progressive Priestly and the French aristocrat Lavoisier were carrying out a sub rosa political duel is also difficult to justify on methodological grounds once comparative progress is laid out. The phlogiston research programme happened to be wrong, but judging that is a cumulative decision starting from its earlier success during the eighteenth century.

Atomic-Kinetic and Phenomenological Theories of Heat

During the last half of the nineteenth-century, there were two principal competing explanations for heat phenomena. One theory is that accepted today: that heat is the motion of small particles making up matter. Or better, there are two theories: that matter consists of small particles or atoms, and that heat is their motion.[25] The second explanation is known as the phenomenological theory because it forbids making specific assumptions about the internal nature of matter and atomic-molecular models specifically. Instead, general descriptive laws relating macroscopically defined quantities—such as temperature, volume, pressure, work performed, viscosity, density, and so on—were taken as the phenomenal appearances to be derived and explained, but suspending judgment on matter's internal structure. The existence of atoms was controversial until after 1900, so the phenomenological theory provided a powerful means for dealing with numerous problems without relying on a molecular reality.

As it turned out, the competition between the atomic and phenomenological theories was mostly the problem of atomic reality, so understanding how the programmes competed and were assessed is a major historical issue. The phenomenological theory has often been associated with modern positivism, as the type of theory abjuring metaphysical "unobservables," and hence, indirectly indicts positivism for its negative influence on the development of the atomic theory. Many have claimed that positivism created an ideological bias that supported the phenomenological theory to the detriment of the kinetic theory.[26] Was there a philosophical perversion of scientific practice through, for example, Ernst Mach or Pierre Duhem, two ardent foes of the atomic theory? Was the kinetic theory unfairly assessed until Einstein's 1905 explanation of random-walk, or so-called Brownian motion, and Jean Perrin's experimental confirmation in 1910?

Peter Clark's research programme analysis argues not. There were first two research programmes—kinetic and phenomenological—with scientists including Rudolph Clausius, James Maxwell, Herman Helmholtz, and Ludwig Boltzmann contributing to *both*. Each programme was remarkably powerful, providing novel quantitative predictions and confirmations, but also, as typical, managing persistent anomalies and observational contradictions. Both programmes were eminently progressive and in aggressive competition for decades. But in the 1890s, the kinetic programme was unable to keep pace and address fundamental problems, in spite of great efforts, while the phenomenological theory continued to progress. Through the work of Boltzmann, the kinetic theory also was clearly in complete contradiction with the phenomenological theory, but Boltzmann's contradicting predictions were believed to have no chance of empirical confirmation. Thus, in the 1890s, the kinetic theory was not just going through a slow period; it was creating exactly the type of consequence of unverifiable effects that provided a raison d'être for the phenomenological approach to begin with.

The kinetic hard core is that heat is "the aggregate of an enormously large number of very small and constantly moving elementary individuals subject to the laws of mechanics" (*MA* 45). Following several heuristic guidelines, traditional mechanical theories then were used to develop a series of molecular gas models, each intended to make improved approximations over its predecessor models. The Lakatosian series of theories involved specific assumptions first on the regular statistical behavior of otherwise chaotically moving molecules; then Maxwell's introduction of a "mean free path," or the average distance a molecule travels before colliding with another; and then additional refinements to assumptions about molecular motion and extensions for specific problems. The kinetic programme was based explicitly on the atomic hypothesis, although with little detail regarding atomic structure or size. Despite this slender basis, the kinetic programme supplied the first reliable estimates of atomic dimensions and thereby considerably strengthened belief in atomic reality.[27] It was precisely such powerful predictions and their confirmations that were taken as the programme's success, notwithstanding the considerable difficulties, and they are representative of Lakatos's idea of progress through a "sea of anomalies."

What of the phenomenological programme? It seems initially odd to even think about a research programme when there are no series of models—that is why it is called the phenomenological theory, after all. The hard core of the phenomenological theory is just that a definite relation always existed between a quantity of heat and the work that in any way could be produced by it, and the aim of the programme is to determine these generalized empirical relations. The programme culminates the phenomenological theory of Josiah Gibbs and Max Planck in the late 1870s, and their fundamental phenomenological laws that in all physical or chemical processes, the total energy of all bodies participating remains constant, and that entropy increases in all physical and chemical processes—in short, the conservation of energy and eventual heat death of the universe. The two phenomenological laws were thought of as all-inclusive empirical generalizations, and the programme's heuristic was to use them to derive new consequences from other empirical regularities. In contrast to the structural models used in the kinetic programme, the phenomenological heuristic was in this way empirically or fact driven: you start with a piece of "experience," like "pressure times volume is proportional to temperature" for gases, and then apply the two laws with specific boundary conditions. For Clark, the phenomenological programme ultimately was limited through its heuristic dependence on available empirical relations since there was no internal mechanism of heat, as there was in the kinetic programme, from which new properties could be theorized. Like the kinetic programme, the phenomenological one was powerfully progressive, and several of its novel predictions could not be derived using the kinetic approach. Helmholtz, for example, following a phenomenological approach, was able to predict in which direction a chemical reaction would proceed and the relative ease of combination or dissociation, or chemical "affinity."

The "experience" of the phenomenological theory was still then theory-laden, but that is consistent with the programme's motivating philosophy, even as expressed in Mach's 1883 historical and critical *Development of Mechanics* and other writings. One of Mach's criticisms of classical mechanics was that Newton's laws were not a priori, as often presented. They were rather, Mach argued, broad generalizations of ordinary human experience with forces and material objects, and the usual perceptions and interactions with them. There is nothing philosophically objectionable about that kind of "positivism," which therefore might "corrupt" physics, no more than Mach's views corrupted Einstein, who found his own use in the special theory of relativity for Mach's critique of Newtonian absolute space and time. The programme's ontological commitments were defined negatively and were unorthodox, but nonetheless led to a research programme whose hard core relied on a coherent conception of "phenomena."

A particular challenge for the kinetic programme was to deduce the basic phenomenological second law of the absolute increase of entropy. Reducing this fundamental law on which so many phenomenological explanations depended to atomic principles was a central challenge for the kinetic programme. One weak interpretation of the second law was that its experimental precision limited the measurement of the reversal of entropy. A stronger interpretation, put forward by Planck, was that the second law expressed a condition of nature and not observation. That is, violations of the second law were not just minute; they never occurred at all. The violation of the second law implied, according to this view, the possibility of a perpetual motion machine, and no exceptions were possible: any single observation of a violation of the second law would reduce the basic content of the second law to nothing.[28]

This second stronger interpretation eventuated in the major conflict between the two programmes. In the early 1870s, Boltzmann apparently had succeeded in providing exactly the needed reduction of the second law to kinetic theory via his so-called *H*-theorem, thus yielding the second law and its consequences as a major prediction for the kinetic approach. But the victory was short-lived, as a critique by Joseph Loschmidt led to Boltzmann's conclusion that the increases in entropy were only statistical, as believed today. That consequence was a direct violation of Planck's stronger interpretation of the second law prohibiting *any* entropy increases. The difficulties posed by Boltzmann's statistical version of the second law, combined with the further inability of the kinetic programme to make substantial progress on other problems, led to a more critical view of the kinetic programme and its allied theory of atoms. In 1900, Lord Kelvin would identify the problems of kinetic theory as one of two great "clouds" over physics, the other cloud being that of the imponderable ether assumed by electromagnetic theory. Problems in heat conduction, viscosity, and thermal diffusion also resisted progress through the kinetic approach. By the end of the nineteenth century, its heuristic power was temporarily exhausted: the research programme was degenerating compared to its phenomenological rival. This does not imply the programme should have been given up, or the "irrationality" of continuing to work on the programme; if anything, its later success shows again that one cannot rule out the possibility that a degenerating programme will recover. But in the 1890s, the programme required what Lakatos calls a "creative shift" in its heuristic approach that would only come after 1900 and the new quantum-theoretic ideas of Planck and Einstein. Between the two programmes there was, as Clark points out, fundamental uncertainty as well as a trade-off between empirical progress and limited heuristic power. The kinetic theory contained a strong heuristic for developing models even though it was the empirically weaker programme, while the phenomenological theory was a consistently progressive and successful programme, with only a weak heuristic for moving beyond macroscopically observable relations.

What was the role of positivist ideology for the kinetic and phenomenological programmes? Technical anomalies of the kinetic programme were acknowledged, such as those associated with specific heats, while the debates involving well-known phenomenologists such as Mach, Wilhelm Ostwald, and Duhem took place on methodological grounds, as a reaction against mechanical models and for empirical-experientially based theories in general. Clark's important claim about this historical interaction between philosophy and science is that "it was the degeneration of the kinetic programme compared with the empirical progress of [phenomenological] thermodynamics which accounts for the rise of scientific positivism," not vice versa.[29] Mach supported the kinetic theory early on, but turned against it when its heuristic power petered out, and not just for so-called philosophical reasons. In contrast, in a kind of ideological "elevation," Duhem included the phenomenological approach as a means for organizing observations in physical theory in general. Lakatos warned exactly against that, meaning "one must never allow a research programme to become a *Weltanschauung,* or a sort of *scientific rigor,* setting itself up as an arbiter between explanation and non-explanation, as mathematical rigor sets itself up as an arbiter between proof and non-proof" (*MSRP* 155). Clark's account shows that the phenomenological research programme should not be held accountable for much of its philosophical ideology.

What largely decided in favor of the kinetic theory was Einstein's theoretical derivation for the possibility of Brownian motion—such as the random motion of a dust mote in a water droplet—including probabilistic equations of particle motion and variability. The effects Einstein described were the result of exactly the deviations from mean thermodynamic quantities predicted by the kinetic theory. As part of his creative shift, Einstein demonstrated in particular that the concept of "average velocity" was not well defined because the probabilistic Brownian paths were almost everywhere "jagged," or nondifferentiable, making certain time averages impossible: the seemingly innocuous assumption of mean velocities turned out to be an invalid "hidden lemma," thus requiring a more refined statistical analysis showing that the *net* movement of particles varied predictably. Though the fluctuations Boltzmann predicted were unobservable, a resulting phenomenon that could be observed, such as Brownian motion, would be a confirmation of atomic reality nonetheless. Numerous attempts were made before 1900 to use Brownian motion—discovered in 1827 by the botanist Robert Brown—to refute the second phenomenological law. These all failed, and Brownian motion did not yet amount to a confirmation for the kinetic theory. Only Einstein's new quantitative analysis, in which he revealed the error of assuming an average particle velocity, plus Perrin's success in experimentally confirming Einstein's calculations, justified Boltzmann's spectacular result that the second phenomenological law was statistical after all, and all this finally stood as positive confirmation of the existence of atoms.

But isn't it "irrational," from a research programme perspective, for the successful phenomenological theory to be cast aside with Perrin's one spectacular result? Remember the strange status of the Gibbs-Planck second phenomenological law: it was a fundamental principle of nature and stronger than a mere research programme modeling axiom. Planck had said in 1883 that the phenomenological second law was all-or-nothing, with "exceptions" making no sense: "The consistent implementation of the second law . . . is incompatible with the assumption of finite atoms. One may anticipate that in the course of the further development of the theory a battle between these two hypotheses will develop *which will cost one of them its life.*"[30] That is just why Boltzmann's *H*-theorem was so controversial. So when Perrin showed that Einstein's Brownian paths behave just as predicted, that was a first major blow to ending the programme. Yet there were also other culminating confirmations: about a dozen new converging calculations for Avogadro's number for the number of molecules in a mole of any substance; and consistently counting and recounting molecules could not be explained on anything but their realistic interpretation. So there were multiple, related confirmations, not a single isolated effect.

Nonetheless, the research programme account becomes tenuous in that the conflict between the two programmes represents a choice between very different conceptions, and even ontologies, for physical theory as a whole: in establishing the reality of atoms, Perrin settled one of the longest and most fundamental controversies in physical theory. As mentioned earlier, the phenomenological theory, with its unorthodox heuristic and antiatomist ontology, is just barely interpretable as a programme. Clark's description is largely correct, but it reaches one limit of research programme historiography. Another example is the special theory of relativity, discussed later, in which the problem of rendering frontier science taxes research programme capacities, one interestingly also motivated by Mach's critical attitude toward mechanical explanations and involving Einstein. In that case, as I will show, there is a clear transgression of research programme categories.

Novelty and Harmony in Copernican Astronomy

The replacement of a geocentric with a heliocentric astronomical system led to far-reaching social changes and controversy, the most notorious being the successful heresy charge against Galileo in 1633 for defending the Copernican system. These changes are not in dispute. The question from a research programme perspective is: to what extent can the Copernican astronomical system be seen, in the last half of the sixteenth century after the publication of Nicolaus Copernicus's *De Revolutionibus Orbium Caelestium* in 1543, the year Copernicus died, as progressive relative to the Ptolemaic? Can acceptance of the Copernican system be justified on internal research programme grounds? Lakatos and his collaborator Elie Zahar wanted to combat Kuhn's view in his 1957 book *The Copernican Revolution* that there are effectively no internal grounds for preferring the Copernican system over the Ptolemaic. For Kuhn, until at least after 1600, the Copernican system was as strong as, but no stronger than Ptolemy's, and therefore its support has to be explained by extrascientific or external factors, at least given a narrow role for what counts as science.[31] That view will not be disputed here. What is of interest is just why the Copernican revolution is poorly represented as research programme progress, but nonetheless helps characterize methodological changes implied by a revised worldview.

Kuhn and Lakatos agree on much about Copernican astronomy. Copernicus's system is, first of all, completely mathematical.[32] Johannes Kepler introduced the idea of explaining planetary orbits on the basis of interplanetary forces, while Copernicus's problem was always to "save the appearances" or account for observed patterns of the planetary system using the geometry of combined circular motions. Copernicus's system does lead to two genuinely novel predictions, the kind that might stand as research programme progress: a detailed description of the phases of Venus, and the apparent motion of stars known as stellar parallax. But neither of these phenomena could be observed with the naked eye, and hence before Galileo turned his telescope skyward. Venus's phases were observed by Galileo in 1609, and stellar parallax by Friedrich Bessel in 1838. So while Copernicus could be said to provide novel predictions, they were not corroborated until after 1600. There were also many astronomical discoveries made in the second half of the sixteenth century and these were often used as justification for the new Copernican view. But the justifications turn out to be odd. In the case of comets and novas, for example, the observations have nothing to do with the earth's motion. A Ptolemaic observer, or an anti-Copernican like Tycho Brahe, could just as easily make and interpret these discoveries, which have no substantive connection to the *De Revolutionibus.* As such, heuristics for planetary dynamics cannot be attributed to Copernicus, nor confirmations of novel predictions. Lakatos agrees as well with Kuhn's analysis that Copernicus's system is neither inherently more accurate nor simpler than Ptolemy's.

With so much agreement with Kuhn's viewpoint, how does Lakatos proceed? Lakatos first identifies the ancient Pythagorean and Platonic astronomical programme whose basic principle was that since heavenly bodies are perfect, all astronomical appearances should be "saved" or described by combining a minimum number of uniform circular motions. The center of the universe could be anywhere, and the programme was defined by an effectively vacuous hard core and the dominating heuristic of uniform circular motion. Ptolemy's innovation with respect to this programme was his introduction of the *equant,* which is a planetary orbit not centered on the earth but on a point offset from the earth's center. Moreover, the rate of rotation of a planet on this circle is not uniform with respect to the earth's center; it is instead uniform with respect to the displaced equant point. Earlier astronomers had already introduced the *epicycle,* which uniformly rotates about the center of yet another rotating circle called the *deferent.* From a methodological perspective, Lakatos says that Ptolemy "violated the [uniform circular motion] heuristic of the Platonic programme and was ad hoc" (*MSRP* 183). Copernicus clearly alleged that his elimination of Ptolemy's equant was an evident advantage and supported the truth of his system. From Lakatos's broad research programme perspective, Copernican astronomy is a return to the Platonic circular motion programme, just using a heliocentric orientation.

The difference between Kuhn and Lakatos's interpretations involves several known, not new, planetary phenomena that are easily explained through Copernicus's system. Since Copernicus provided no novel predictions confirmed until after 1600, for a research programme analysis to "justify" the progress made by Copernicus, such novel deductions of known facts are essential. This slight change in characterizing research programme progress was suggested by Zahar and accepted by Lakatos as part of their joint study of the Copernican programme: a known fact can count as novel relative to a programme if the programme was not "rigged" to account for the fact but offers a qualitatively new explanation. The appearances identified as having novel Copernican explanations in this revised sense are that: planets have stations and retrogressions; the periods of superior planets as seen from Earth are not fixed; the apparent motion of the planets, when described from Earth as fixed, will include as a component the motion of the Sun; and the elongation of orbits of inferior planets is bounded, with the periods of the planets increasing with their distances from the Sun. The first three observations are almost immediate consequences of Copernicus's system and heliocentric geometry. The last can be calculated using an argument that, in Ptolemy's system, has to be replaced by ad hoc assumptions regarding the radii of epicycles to deferents. Copernicus's "excess predictive power" comes, for Lakatos, from a better portrayal of this one set of known appearances.

Kuhn's account of these same derivations is simply Copernicus's own account. Copernicus framed his novel explanation of known facts in terms of what he described as the harmony, naturalness, and coherence demonstrated for his approach. Copernicus even acknowledged that the relevant astronomical appearances can be explained by either the Ptolemaic or Copernican system, but then points out the harmony of the more direct Copernican explanation. As Kuhn writes:

> There are [a] great many such arguments. The sum of the evidence drawn from harmony is nothing if not impressive. But it may well be nothing. "Harmony" seems a strange basis on which to argue for the earth's motion, particularly since the harmony is so obscured by the complex multitude of circles that make up the full Copernican system.[33]

Harmony, for Kuhn, methodologically amounts to *nothing* in the context of Copernicus's other weaknesses vis-à-vis calendrical calculation and quantitative precision, and is largely a Platonic-ideological prop adding no predictive power or accuracy.

How important, therefore, can harmony become when dressed up as Lakatos's novelty? Extending physical-dynamic arguments to astronomy would be a major innovation of Kepler and Galileo, and interpreting Copernicus's mathematics using types of assertions that would later become central to astronomical theory is anachronistic. What differentiates Kuhn and Lakatos's accounts are not historical facts but a slender difference in the methodological content of Copernicus's mathematical harmony. If Copernicus had a substantive dynamic theory of planetary motion, from which his predictions were derived, one might plausibly consider harmony to express the more powerful heuristic of such an approach. But Copernicus's system is strictly mathematical. Even Brahe constructed his own geocentric system mathematically equivalent to Copernicus's; at the time, there was no further physical theory to be integrated into astronomical calculations. The common goal is to save the planetary appearances, and so greater elegance in mathematical derivations has little methodological advantage, contrary to what Lakatos might contend. Kuhn's dismissal of harmony as Copernican rhetoric or Neoplatonist ideology is methodologically correct, and conversely, Lakatos and Zahar's repackaging of Copernicus's words as if they are indicative of greater explanatory power is difficult to maintain.

The problem is that although Lakatos describes even Ptolemy's work as part of an ancient Platonic astronomical programme, Copernicus's work occurs at the beginning of modern science and the emergence of research programmes on a larger historical scale. Copernicus has one foot firmly in the Aristotelian project of saving the phenomena, and the other barely in a new science of correcting the senses and generating novel predictions. Lakatos makes a virtue out of expanding the horizon of facts accounted for by a programme, but that is a modern bias. What for Lakatos is "degeneration" or stagnation in failing to produce new facts, can be saving appearances in the medieval Aristotelian view in which the universe of facts, so to speak, is finite. There is not an arbitrarily large inventory of facts to explain, only those accessible mainly through ordinary perceptual resources, like naked-eye observations of the stars before Galileo. Novel explanations, then, cannot advance where a boundary exists instead of a frontier.

If that is the case, then as Feyerabend points out, "The revolution that was started by Copernicus's hypothesis was much more dramatic than a transition, *within the methodology of research programmes,* from one research programme to another. It brought in new standards and thus constituted a true paradigm change in the sense of Kuhn."[34] The difference lay in the transition from what Feyerabend calls a cosmological assumption of an Aristotelian "finite" world of appearances to be explained, with no need of constantly improving research programmes, to one in which hidden realities are taken for granted and the horizon of facts may constantly expand. Galileo perhaps more than anyone inaugurated that new epistemology, and Lakatos's research programme account of Copernicus founders on just this historical change. If indeed Copernicus was one of the last to work within the old Aristotelian methodology, then that is a more decisive methodological characterization than an appraisal that ignores a world-historical change in methodological standards. More than that, implicitly relegating Copernicus's Neoplatonism to "external" history now becomes a historical falsification of its own. Call it what you will, either external history or a history of changing methodological standards is needed to understand the Copernican revolution. The methodology of scientific research programmes is but one style, albeit a significant one, of scientific reason, and it does not fit everywhere.

Heuristic Power in the Special Theory of Relativity

Einstein's special theory of relativity (STR), published in 1905, provided a new starting point for physics by proposing a unified view for the roles of motion and the propagation of light and electromagnetic waves. The major innovation of this approach was that Newtonian absolute space and time were given up for space-time coordinates defined relative to a given reference frame, without any one frame being privileged over another. The STR implied the dependence of an object's mass on its velocity, the absolute limiting velocity of the speed of light, and the "covariance" of electromagnetic laws relative to inertial, or nonaccelerating, reference frames. Covariance means that while physical measurements, say the velocities of moving bodies, may differ across reference frames, these quantities obey the same electrodynamic laws regardless of the chosen reference frame. The STR does not address gravitational forces and accelerated reference frames, which are part of the general theory of relativity published in 1915. As well, the STR is famous as ending the need for an ether through which light and electromagnetic waves could propagate. The STR, therefore, accounted for the effectively null results of Michelson and Morley in the 1880s in attempting to measure differences in the speed of light relative to the earth's motion through the ether.[35] These striking consequences were founded on two principles characterizing the STR hard core: (*a*) Maxwell's electromagnetic laws should not vary with relative motion; and (*b*) the velocity of light is constant as measured in all reference frames.

A remarkable historical aspect of the changes brought about by STR is that many of the theoretical tools used in the theory were formulated within earlier Newtonian and ether-based physical theories of Hendrik Lorentz. Until 1905, Lorentz was esteemed as the leading theoretical physicist of the time, with his electron and ether theory of electromagnetism recognized as the main approach integrating optical, electrostatic, and electrodynamic phenomena. Lorentz's work was also at the center of attempts to build physics, including theories of mass, on an electromagnetic foundation. The electron theory was the last great theory taking Newtonian concepts of absolute space and time to their limits, to the extent that many of the individual changes introduced by Einstein had been introduced by Lorentz as a means to reorient physical theory around electrodynamic fundamentals.

One way of making the comparison is that Einstein provided a framework in which several mathematical ideas that, for Lorentz, had unclear or ambiguous physical interpretations in Newtonian space-time could be given unambiguous realist interpretations from a relativistic perspective. The notion of a "local" or relative time for moving reference frames appears in Lorentz, as does the Lorentz-Fitzgerald contraction, or multiplicative "shrinking" factor $\sqrt{(1 - v^2/c^2)}$, representing the contraction of a moving object; v is the velocity of a moving body and c the speed of light, which Lorentz, like Einstein, made a limiting velocity. The general "Lorentz transformations," used to establish the covariance of electrodynamic laws for different reference frames, were also developed by Lorentz by 1904 yet rediscovered by Einstein. Like Einstein, Lorentz was motivated less by particular results, such as the Michelson-Morley experiments, than by the goal of unifying mechanics and electromagnetism, seen by many as two mostly disconnected branches of physical theory. Mechanical explanations in physics were criticized by Mach and Heinrich Hertz, among others, and one approach to unification was the development of a physical theory based on electrons and electromagnetic principles alone. This meant, for example, that matter could ultimately be reduced to electrons, and since the laws of electrodynamics and the ether should govern the motion of electrons, classical mechanics, too, ultimately might rest on an electromagnetic foundation. Lorentz did not completely embrace such an electromagnetic worldview but his theories came close to succeeding.[36]

In this way, Lorentz's programme was the one most clearly in competition with Einstein's STR. The question for Lakatos's philosophy of science is the extent to which the decline of Lorentz's programme and the rise of Einstein's can be characterized via the methodology of scientific research programmes. Elie Zahar depicts Lorentz's "ether programme" through a hard core consisting of Maxwell's laws of electromagnetism, Newton's laws of motion governing particles moving through the ether, and the properties of an additional force—sometimes called the "Lorentz force"—that mediates the action of particles and the ether. Lorentz's ether occupied the same space as charged particles and permeated uncharged matter, and was not moved by bodies as they moved through it. The ether was not linked mechanically to "ponderable" matter; rather, the Lorentz force accounted for electrostatic and electrodynamic relations between matter and the ether via electrons. The theory started out as only an electron theory, but came to include the physics of all matter, not just charged matter. In traditional mechanics, masses, forces, and the shapes of bodies could be specified independently of their motion, while in Lorentz's electron theory, they would depend on their motion through the ether, and their properties ultimately depended on the ether's electromagnetic properties.[37] Lorentz's hard core is complemented, then, by what Zahar characterizes as the programme's heuristic: that "all physical phenomena are governed by actions transmitted by the ether."[38] As such, the programme combined the mechanics of particles and fields of electromagnetism, both thought of in a Newtonian world of absolute space and time. It was this dual basis that led Lorentz to several of his innovations, always within classical physics. Mechanics and electromagnetic theory were both used, but were effectively separate sets of theoretical principles. One notable conclusion was that since the ether acted on particles and not vice versa, Newton's law of action and reaction was not absolutely valid. This consequence was not taken as an anomaly but a consequence of the limitations of mechanics from the perspective of electron theory—that for electron-ether interactions, this "aggregate" law, valid at the macroscopic level, may not apply.[39]

From this perspective, Zahar claims first that Lorentz's programme was theoretically and empirically progressive. This assertion is not contentious. By 1904, when Lorentz had largely developed the final version of his electron theory, he had been able to explain scores of problems including the Doppler effect and Zeeman effect, named for Lorentz's student with whom he shared the 1902 Nobel Prize. The major anomaly for the programme would be the same as that posed for all turn-of-the-nineteenth-century physics: the discovery of Planck's energy quantum. Lorentz was among the first physicists to deal with quantum issues for his own theory, finding that his prediction for black-body radiation spectra had the same limited validity for long wavelengths as did other radiation theories, and that the quantum hypothesis would imply major changes in the electron theory. Among the problems Lorentz addressed was the null result of the Michelson-Morley experiments. Michelson's interferometer apparatus was accurate enough to detect velocity differences for light on the order of 10^{-8}, yet none was observed. It was through this experiment that Lorentz in 1892, and also George Fitzgerald, proposed an explanation of the Michelson-Morley null result based on the compensating minute contraction of Michelson's interferometer arms by the factor $\sqrt{(1 - v^2/c^2)}$. Lorentz's arguments were recognized at the time as ad hoc, but his overall progress was so powerful that it matters little in comparison with the STR.[40]

Leaving such possible "demerits" of Lorentz's programme aside, how do Einstein and Lorentz's compare in terms of their comparative predictive content? Tests conducted by Walter Kaufmann and Alfred Bucherer, in 1905 and 1909 respectively, weakly confirmed and disconfirmed each of Lorentz and Einstein's theories, so from these results, too, the comparison of the two programmes is roughly "equal." Lorentz's electron theory and the STR were also shown by Henri Poincaré to be observationally equivalent. No experiment could distinguish between their consequences, and after 1905, the ideas of contraction and local time were often combined under labels including the "Lorentz-Einstein," "relativity," or "Lorentz" theories. Zahar wants to give positive reasons in terms of research programme progress of not necessarily why Lorentz's programme should have been given up but why, given its recognized power and success, and given the observational equivalence with the novel and unorthodox STR, anyone should see a comparative research programme benefit in pursuing the STR. Given Poincaré's proof of the observational equivalence of the two programmes, the historical and methodological challenge, largely self-imposed by assuming the research programme apparatus, is considerable.

Zahar's answer begins with his characterization of Einstein's STR programme. Both the relativity and light principles, as stated by Einstein and summarized by (*a*) and (*b*) above, are implemented, says Zahar, according to two primary heuristics for the deduction of new physical laws: "[*i*] a new law should be Lorentz-covariant and [*ii*] it should yield some classical law as a limiting case" (*MA* 256). "Limiting case" means, for example, that when velocity/speed of light = v/c approaches zero in a relativistic law, meaning that velocity is small compared to the speed of light, then a classical law should result. The first part of the heuristic indicates how to generate new laws, such as a relativistic version of Newton's second law of motion: one looks for the reformulation that will be Lorentz-covariant under relative motion. That truly is a clear and powerful heuristic; the mathematical technique is explicit and can be directed at existing classical laws of motion and electromagnetics. As Zahar points out, Planck's modification of Newton's second law, $f = ma$, provided an early application of this heuristic. Similarly, toward the end of 1905, Einstein discovered his equivalence of mass and energy in the formula $E = mc^2$, generated by application of the heuristic to developing a relativistic version of the conservation of energy, or one holding in all inertial reference frames. The result was another contribution to the unification of mechanics and electromagnetism, and one that Einstein considered to be the most important consequence of his electrodynamics.[41] Still, no truly novel predictive consequences yet followed, and Lorentz and others could at least temporarily see the STR as just another approach to achieving the shared end of unifying mechanics and electrodynamics.

Zahar's answer is that the comparison between the programmes has to be based solely on their heuristic power. He effectively backs into this historical explanation from his claim that the electron programme was progressing, as well as the lack of distinguishing predictions between the two programmes. The heuristic of Lorentz's programme, though having addressed the relation of ether to matter, and explaining most known optical and electromagnetic phenomena, was "running out of steam," while Einstein's provided a new approach to unifying mechanics and electromagnetism.[42] This heuristic power is expressed through the derivation of new relativistic laws, like

Planck and Einstein's. For additional empirical progress within the relativity theory, Zahar maintains that it is necessary to look to the 1915 *general* theory of relativity incorporating gravitation and predictions such as the known slight precession of Mercury's orbit, recognized as an anomaly in Newtonian astronomy. If the STR has a research programme advantage over the electron theory circa 1905, it is in its sheer heuristic, but not predictive, power.[43]

But only by a considerable liberalization of research programme progress criteria does the STR count as "progress" over Lorentz's electron and ether programme. Perhaps for Hermann Minkowski or Planck, with the talents to explore Einstein's heuristics, the opportunity was evident, but that perception was far from shared, and it took many years for the new approach to be understood. The deficiency in the research programme account is that while Lorentz's work is well described as a research programme with a hard core, ether-electromagnetic-mechanical heuristics and a succession of theories incorporating progressive versions of the Lorentz-Fitzgerald contraction, Einstein's programme is based on a new universal principle and not the design of explanatory models. The relativity principle is effectively a metaprinciple about the desired covariance of physical laws under relative motion, and indeed is used in the first instance to revise physical laws, not theories making use of those laws. Einstein himself was aware of this crucial distinction, especially after having tried to proceed by a direct or "constructive" model-building approach. He compared his approach to that of phenomenological thermodynamics as founded on its two laws, which were empirically generalized principles taken as universally true, but that did not direct the development of models. The phenomenological theory of heat, as we saw, is still sufficiently "inside" physics so that it can be reasonably construed as a research programme. But when all physical laws are up for reformulation—as are notions of mass, space, and time—there is no framework that defines a research programme context. In this way, the research programme approach fails to capture Einstein's revolution, but does so "from below," by identifying how Lorentz's ideas were artfully redeployed through Einstein's heuristics. Both the kinetic theory of gases and Lorentz's theories represented traditional, constructive, model-building approaches, and part of Einstein's innovation was to define relativity as a universal principle, thereby overcoming the implicit dualism of separate, classical mechanical and electromagnetic explanations as used by Lorentz. The STR was not a closed theoretical system, and Einstein repeated this view in his autobiographical writings decades later.[44]

Lorentz's research programme was a climax of classical physics, and Einstein's approach implied abandoning absolute motion, time, and space, which Lorentz still assumed. More than that, neither mechanics nor electromagnetic theory were to be privileged by Einstein. Physical laws, like relativistic conservation of momentum or Einstein's new mass-energy conservation, would link concepts from *both* in novel ways. The relativity principle was the heuristic key to accomplishing both these tasks, but it goes beyond garden-variety research programmes. As put by historian Tetu Hirosige, "Einstein was seeking to bring about a unification of mechanics and electromagnetism with regard to the relativity of motion. He had set himself a problem concerned with the form, rather than the content, of theory."[45] Such fundamental heuristics of "form" are rare and the stuff of the greatest conceptual novelty. In this case, the novelty goes beyond that which can be accounted for by research programmes alone. Lorentz, like Ptolemy compared to Copernicus, had prima facie as powerful a theory as did Einstein circa 1905.

The treasures of the new approach were not apparent at the time of their introduction and are poorly accounted for by research programme categories, even with considerable hindsight. For both Copernican astronomy and the STR, relative to their competitors, science is changing as much in methodological innovation as physical theory. At best, the research programme analysis helps identify just what those combined methodological-physical innovations were. Einstein, in particular, introduced an entirely novel theory of physical measurement through his analysis of reference frames and event simultaneity. Einstein himself recognized the need for flexibility in revising received physical concepts and how this creates problems for the reconstruction of science. "The external conditions," remarks Einstein, "which are set [for the scientist] by the facts of experience do not permit him to let himself to be too much restricted, in the construction of his conceptual world, by the adherence to an epistemological system. He, therefore, must appear to the systematic epistemologist as a type of unscrupulous opportunist."[46] Unscrupulous? Not *completely* from a research programme perspective, and less than one might expect, but definitely unique and transcending the comparative power of Lakatos's methodology.

8 | Contradiction and Hindsight

Summarized abstractly, it is not clear why Popper would find Lakatos's approach as objectionable as he did.[1] The research programme approach encourages an intimate understanding of science that nonetheless helps make science exoteric. The rules respect the flexibility of informal science and provide some critical standards that, while not binding, are of evident historical and even practical use. Public policy debates involving, for example, environmental research programmes in climate change or electromagnetic field exposure risk might be cast in terms of the hard cores of competing programmes and their predictions, corroborations, anomalies, and progress or degeneration. The logic of research programmes is not obscure and respects the limitations of human judgment.

Popper's engine is falsification, and it is fueled by contradictions between theory and theory-laden observations. As Popper never tired of pointing out, you cannot logically prove a theory but you can falsify one via a refuting observation or designed crucial experiment. That logical asymmetry and the role for contradictions with experience as the indicator of theory falsehood is the basis of his entire approach. It was natural, then, that Popper maintained that consistency "can be regarded as the first of the requirements to be satisfied by every theoretical system, be it empirical or non-empirical" (*LSCD* 92).[2] As well, somewhat hyperbolically, he believed that "the acceptance of contradictions must lead . . . to the end of criticism, and thus to the collapse of science."[3] Noncontradiction for Popper is what Kant would consider a "regulative" principle, meaning that it delimits the range and types of legitimate inferences made by human reason.[4] Popper assumes no foundation for knowledge in the senses or reason, and accepts that refuting observations are themselves theory-laden, not positivistic hard facts. But since progress is governed by refutations ultimately grounded in logic, noncontradiction retains a privileged role.

A theory's "potential falsifiers" are the conditions that, were they to be observed, would refute it, and a theory is "scientific" for Popper if it has some potential falsifiers. For refutations to be effective in eliminating false theories, Popper wants potential falsifiers to be specified in advance: "Criteria of refutation have to be laid down beforehand: it must be agreed which observable situations, if actually observed, mean that the theory is refuted."[5] Popper's motivation, which is laudable, is to exclude post hoc rationalizations of the type he found in vulgar Marxism and the Freudian theory of the 1920s. One should not decide *later* what "counts" as a crucial experiment or refutation, since then contradictory or anomalous results might be explained away by ad hoc adjustments. So, in addition to identifying conditions of refutation in advance, falsifications or crucial experiments should be definitive. For given that observations are theory-laden, if falsifications are not somehow "final," then reinterpreted observations may allow theories to move in and out of "falsified" status, rendering any systematic progress incoherent. It is also a mark of scientific *dishonesty* not to abandon a theory whose falsifying conditions have been observed.

Lakatos rejects almost every part of Popper's conception of falsification. The spirit and letter of Lakatos's research programmes embrace a positive role for contradictions and backward-looking, retrospective reinterpretations of experiments. Lakatos devises his own characterization of what is scientifically ad hoc and dishonest, but they are not Popper's conditions. Contradictions or anomalies—just inconsistencies between theory consequence and observation—are ubiquitous in science, Lakatos argues. "All theories are born refuted," Lakatos liked to say; they abound in inconsistencies and anomalies—they are always already falsified as it were. In addition, many research programmes—including Newtonian mechanics, Maxwell's electromagnetic theory, or relativity or evolutionary theory—also could never have identified conditions in advance that, if observed, would justify giving up the entire research programme. To the contrary, Lakatos claims, some crucial experiments are seen as falsifications only in hindsight as they are reinterpreted against a changing landscape of competing programmes and evolving theories. Lakatos's role for hindsight introduces a historical role within science itself and not just the reconstructions of the historian or philosopher.

Lakatos's role for contradictions as ubiquitous also offers a modern interpretation of Hegel's insistence that contradictions are inherent in all concepts, with a principle difference being that our "unit" of analysis has shifted from quasi-cognitive concepts to articulated theories. Popper, therefore, plays a role analogous to Kant, in that for Hegel, Kant's

> failure to make a more thorough study of Antinomy was one of the reasons why Kant enumerated only four Antinomies. . . . [T]hey appear

> in all objects of every kind, in all conceptions, notions, and Ideas. . . . In modern times it was, more than any other, Kant who resuscitated the name of Dialectic, and restored it to its post of honor. He did it, as we have seen, by working out the Antinomies of reason.[6]

Kant, for Hegel, introduced antinomies as a central philosophical category, but unrightfully limited their positive role. So too, Popper introduced falsification, and Lakatos contends that contradictions cannot be limited within the boundaries Popper desired. Kuhn and Feyerabend also saw that science progressed happily in a "sea of anomalies." Lakatos's Hegelian difference is to take scientists' ability to reason through contradictions as a shared and largely objective feature of scientific interpretation and rationality.

Lakatos's principal evidence for the positive role of contradictions in research programmes, against Popperian crucial experiments and for *post festum* historical interpretation or hindsight, is ultimately historical. Not that history per se is privileged, but that the richness of decision-making practices drawn from great science demonstrate the rationale for both contradiction and hindsight, while also empirically "refuting" Popper. Lakatos makes extensive use of several famous episodes from modern physics, principally Niels Bohr's early quantum model of the atom, to argue for "progress on inconsistent foundations." For hindsight, Lakatos retells the Michelson-Morley experiments on the speed of light, which Popper and many others described as a great crucial experiment refuting classical Newtonian physics and leading to the relativity theory.

The Old Quantum Theory

Lakatos reconstructs Bohr's early quantum theory as a research programme based on a hard core consisting of Bohr's five postulates for his atomic model, articulated clearly by Bohr in an epochal 1913 paper. Bohr's famous idea was that electrons could travel only in orbits of discrete, not continuous, radii, and that within an orbit, an electron does not radiate energy. Instead, the electron radiates as it "jumps" between orbits, and then the quantity of radiation is not continuous but a quantum multiple; thus did Bohr integrate Planck's energy quantum into atomic models. Models based on Bohr's programme enjoyed several successes in predicting wavelengths of emission spectra, especially in explaining their discrete light-and-shadow character, until the mid-1920s when the programme was superseded by Louis de Broglie's wave mechanics.[7] Bohr's programme developed, according to Lakatos, as its hard core was elaborated through a sequence of theories: first, a model of the hydrogen atom with the electron moving in a circular orbit;

then a fixed-plane elliptical orbit; then removal of the restriction that the orbital plane and nucleus must be fixed, and so on. The successive models each explained a new or existing spectral series while simultaneously being confronted with anomalous results that were reinterpreted using a later model.

One role for contradictions in Bohr's programme is that its hard core was inconsistent with the then well-corroborated Maxwell-Lorentz electromagnetic theory predicting that Ernest Rutherford's, and then Bohr's, atomic models were inherently unstable because orbiting electrons should continuously spiral into the atomic nucleus. As well, Bohr was faced with contradictory experimental observations, primarily the spectra of elements he could not explain. Bohr had no plan for overthrowing the electromagnetic theory with which his programme conflicted, so his progress occurred in the context of significant theoretical and observational contradictions. He indicated the need for electromagnetic theory to be replaced eventually, yet the "refutable variants" of Bohr's hard core, the "series of theories" that the programme generated, were still inconsistent with the Maxwell-Lorentz theory. As Lakatos says, Bohr's "positive heuristic, even if it had been completely successful, would have left the inconsistency with the Maxwell-Lorentz theory unresolved" (*FM* 56). The history of the programme, for Lakatos, shows that one can countenance "growth-on inconsistent foundations" (ibid.); "how a progressive shift may lend credibility—and a rationale—to an inconsistent program" (*FM* 64); how there can be "marvelously fast progress—on inconsistent foundations" (*FM* 67); and that "some of the most important research programs in the history of science were grafted on to older programs with which they were blatantly inconsistent" (*FM* 56).[8] Such liberalism is not a "crime against reason," as Popper would have it, but a means for the rational scientist to exploit the heuristic power of the grafted programme "without resigning [one]self to the fundamental chaos on which it is growing" (*FM* 90). No special extrascientific explanation is needed, for Lakatos, to account for such progress.

At first blush, Bohr's early quantum theory looks like an ordinary research programme. As Lakatos shows, however, it is distinctive in epitomizing the role of contradictions and anomalies. Bohr's old quantum theory, as it is quaintly known, occurred within, and was a response to, a situation in physics in which the contradictions between the classical view of continuous energy and the new idea of a discrete quantum of energy introduced by Planck in 1900 were among the most extreme, widespread, and penetrating in the history of physical theory. Classical physics was generally seen by about 1910 as incapable of accounting for atomic or molecular processes, but nobody saw how classical models could be modified. It was not just a question of getting the right model from the classical view. *Any* dynamic equilibrium of moving electrons would dissipate energy and lead the system to collapse. In order to define a stable atomic model, Bohr made an assumption that was in direct and complete contradiction to Newton's mechanics and Maxwell's electrodynamics: namely, that electrons have a discrete set of permissible stationary orbits, and as long as the electron remains in a stationary orbit, no energy is radiated.[9] Completely contrary to classical physics, Bohr said an electron may jump at random times and instantaneously from one orbit to another of lower energy, with the lost energy emitted as electromagnetic radiation of a specific frequency.

Bohr also did not present his theory as a completed or realistically viable alternative. It was, as he described it, merely a "preliminary and hypothetical" way of representing results that defied all classical mechanical and electrodynamic explanation. In an address delivered before the Physical Society in Copenhagen on December 20, 1913, Bohr said his goal was not "to propose an explanation of the spectral [that is, emission] laws" but rather "to indicate a way in which it appears possible to bring the spectral laws into close connection with other properties of the elements, which appear to be equally inexplicable on the basis of the present state of the science."[10] As historian Max Jammer puts it, Bohr's intention was not to give a satisfactory answer to a definite question but to search for the right question to ask. Bohr saw the deficiencies of his atomic model far in advance, yet he was convinced that progress would only be achieved by exposing the antitheses between the quantum and classical views as deeply as possible. Nonetheless, Bohr's approach, based on his quantum orbit postulates, follows largely orthodox research programme directives.

Dealing with the contradictions of classical physics introduced by quantum concepts occurred all across physics at the time. But Bohr's approach was different from that of Planck and Einstein in that he was not trying to build a bridge between old and new physical views. His conceptual scheme represented a significant discontinuity with classical physics. Bohr did make use of what he would later call the "correspondence principle," by which a classical physical description could be derived when the scale of events is sufficiently large to neglect individual quanta, and the probabilistic account of quantum phenomena might be taken as a generalization of classical descriptions. That linkage, though, was not Bohr's intent. While using classical concepts, say of momentum, to facilitate the interpretation of results, he also made it explicit that no classical mechanical explanation was intended. In this way, Bohr used his atomic models to work both sides of a contradictory quantum divide.

For several years following the appearance of Bohr's 1913 theory, some thought the old quantum theory could provide the basic approach to atomic modeling. Arnold Sommerfeld made perhaps the greatest contributions using an orthodox heuristic, in which Bohr's models were taken at face value, instead of as a stepping stone to the next big idea. What was needed, from this perspective, was principally more experimental work and refinement, in contrast to Bohr's provisional use of his programme to get to the next stage of theoretical development. Sommerfeld, for example, used elliptical electron orbits for the analysis of the fine structure of the spectrum of the hydrogen atom. But for Sommerfeld, the correspondence principle really was "a magic wand that allowed the results of the classical [electromagnetic] wave theory to be used for the quantum theory," which Bohr held as he marched forward with what was ultimately a conceptually incoherent blend of quantum and classical ideas, intended as the bridgehead to an as yet unknown approach.[11] Sommerfeld saw that Bohr was using the correspondence principle—say, by letting Planck's constant h tend to zero—to derive results he could not achieve otherwise, and in this sense, the "pure" quantum theory was incomplete. According to Bohr's student Hendrik Kramers, while the correspondence principle was definitely a part of the quantum approach, both the classical and quantum views were false "caricatures" of reality, with neither one being true.[12]

In this way, Sommerfeld and Bohr used differing heuristic approaches to make use of the shared hard core originating in Bohr's 1913 approach. Actively promoting contradictions for Bohr was just part of how the programme, even taken as Kramer's caricature, would progress and then be superseded, which it was, showing Sommerfeld's hope for an orthodox quantum theory to be misplaced. The programme's clearly articulated structure gave Bohr's revolutionary admixture of truth and falsity both stability and a means of directing progress. The demise of the old quantum theory was one of diminishing research programme progress as well. Bohr's atom was a one-electron model, and all methods failed to extend the approach to helium, the second smallest atom, not to mention heavier ones. Bohr's account of the structure of the periodic table was ad hoc, and phenomena such as the so-called anomalous Zeeman effect resisted progress, too. Bohr's approach did not motivate a role for Einstein's new photons, which Bohr made no use of, while experiments such as Arthur Compton's X-ray scattering in 1923 were making a role for photons essential.

With his research programme account of Bohr's early work, Lakatos also lays the foundation for his ideas on the changing interpretations of experiments by arguing that the confirming role of observations for Bohr's programme depended strongly on the sequence of atomic models and spectral observations as they appeared historically. For example, in 1913, Bohr explained the Swiss schoolmaster Johann Balmer's 1885 formula for the hydrogen spectrum. While Bohr's theory quickly predicted the structure of additional spectra, the deduction of the Balmer formula is not quite a novel prediction in that the spectral series was known and Balmer's formula existed. Had Bohr's quantum theory been used by Balmer, Bohr's derivation would "be seen not as a mere theoretical reinterpretation of [Balmer's formula] but as a new fact in its own right" (*FM* 70). But Balmer's formula *was* known before Bohr's models, and therefore, the latter could be taken as a "mere reinterpretation" and not a bold prediction of the spectral pattern; this kind of skeptical challenge was one that Bohr faced repeatedly in defending his atomic model.

For Lakatos, this shows that historical phasing, or the sequence in which theoretical explanations or predictions and their corroborations or refutations appear, is crucial for the interpretation of progress, even as no logical conditions have changed. In general, "the interaction between the development of the programme and the empirical checks may be very varied—which pattern is actually realized depends only on historical accident" (*FM* 65). In Bohr's case, his theoretical account contrasted sharply with the purely algebraic formula of the numerologist Balmer, who found regular patterns in all kinds of natural phenomena. But to continue the example, had more corroborating spectral series been produced by a rival programme prior to Bohr's work, then a series of models explaining them, like Bohr's, "will not represent an empirically progressive problemshift and therefore, although all the evidence supports his theories, the scientist has to work on further in order to prove the scientific value of his programme" (*FM* 66). Or imagine an isolated theoretician working out models until finding that programme's most sophisticated refutable variant. If corroborating spectral series arrive after the appearance of such a theory, the results may be interpreted as novel predictions. "The theoretician—at his desk—is here seen to work far ahead of the experimenter: we have a period of relative autonomy of theoretical progress" (ibid.)

The alteration of refutations and confirmations found in Bohr's actual programme lay between these two extremes and created some of his greatest challenges. Bohr faced what at first appeared to be clearly contradicting spectra, notably those for ionized helium found by Charles Pickering in 1896 in starlight, and again by Alfred Fowler in 1912 in the laboratory. A single ionized helium atom is a one-electron atom, and Bohr's model should thus apply. Bohr rederived calculations for the Pickering series and suggested an experiment that turned out to support his claim that ionized helium, not hydrogen, was being observed. Fowler then argued that the newly predicted lines, while an improvement, were not close enough. By changing his mass calculation, Bohr found that he could exactly account for Fowler's objection, and this particular spectral explanation was one of Bohr's most dramatic triumphs.

Lakatos's historicist lesson is that varied realizations of the same set of "unordered" theories and observations can lead "rationally" to differing depictions of progress. The notion of phase differences in history is also one found in Hegel and Marx, the latter formulating it as the "uneven development" of economic formations in history, and Hegel using it to describe similar patterns in the development of the nation-state in his work *Reason in History,* the introduction to the *Lectures on the Philosophy of World History.* There clearly is no "inevitability" to historical development in Lakatos; neither is there any Archimedean point from which to assess the "right" interpretation of progress, which depends not just on a research programme's current state but how it got there. A normative consequence, for Lakatos, of the role of phase differences in scientific progress is the need for tolerance in evaluating research programmes. Bohr's example vis-à-vis the Balmer formula suggests that budding research programmes may begin by producing novel explanations of old facts and a new programme should be sheltered from its rivals as long as it can be progressively reconstructed; moreover, "we should certainly regard a newly interpreted fact as a new fact, ignoring the insolent claims of amateur fact collectors [like Balmer]" (*FM* 71).[13]

Contradictions in Theories and Nature

For Lakatos, the critical issue in how contradictions become refutations is that a problematic observation or fact does not determine which theory it refutes. Theories are always coordinated against a large background of undisputed knowledge—say, instrumentation, observational theories, or accepted approaches to a range of specialized problems. Schematically, if P_1 & P_2 & . . . & $P_n \rightarrow Q$, but not-Q is "observed," then many subsets of the P's might be taken as false on logical grounds, and in practice there often are several realistic alternatives to choose from. Lakatos, therefore, promotes a holism similar to that associated with Pierre Duhem and W. V. O. Quine, the latter long-ago famously asserting that even logic or mathematics was open, if you wanted, to empirical refutation. Where Lakatos differs from Quine is that Quine provides no further account of how refutations are "directed." Lakatos even sees Quine's holism as a concession to an irrationalism characteristic of pragmatism, in that almost any choice can be deemed pragmatic. Popper also, for Lakatos, provides an equally undesirable answer—namely, to sharply delimit one's "potential falsifiers," "accepted" ceteris paribus conditions, and finally a "theory under test," since that extensive stage setting will not be possible in practice. But it is often the simultaneous redefinition of all these boundaries that makes up a refutation, not just a falsifying observation falling into a place set in advance. Lakatos's alternative description of these interactions and negotiations of truth is as a pluralistic "clash" between observation statements and competing explanatory theories:

> The problem should not be put in terms of whether a "refutation" is real or not. The problem is how to repair an inconsistency between the "explanatory theory" under test and the—explicit or hidden—"interpretative" theories; or, if you wish, the problem is which theory to consider as the interpretative one which provides the "hard" facts and which the explanatory one which "tentatively" explains them. In a monotheoretical model we regard the higher level theory as an explanatory theory to be judged by the "facts" delivered from outside (by the authoritative experimentalist): in the case of a clash we reject the explanation. In a pluralistic model we may decide, alternatively, to regard the higher-level theory as an interpretative theory to judge the "facts" delivered from outside: in case of a clash we may reject the "facts" as "monsters." In a pluralistic model of testing, several theories—more or less deductively organized—are soldered together. (*FM* 44)

Such, in brief, is Lakatos's portrayal of how contradictions are tolerated and managed. For the purpose of following the role of contradictions in Lakatos's thinking, it is necessary to recognize the change to the modern idiom of statements and theories from Hegel's fairly unpalatable idea of contradictions in concepts and *things.* The Anglo-American philosophical tradition, inspired by Frege, Russell, and modern logic, often uses formalized theories as models of scientific ones. Even if the actual translation into a formalism is problematic, the locution "theory *T*" refers to some idealized formal representative that, in principle, can be analyzed using conventional logical methods of syntactic or semantic analysis. Popper, Kuhn, Lakatos, and Feyerabend, along with others following after them, rejected the general need or desirability of formalizing scientific discourse for philosophical analysis. The classical law of noncontradiction, of "not (p and not-p)," is a standard consequence of most formalized theories, and so if contradictions are ubiquitous in scientific theories, it is not obvious how they are to be formally translated. After the fact, some translation and segregation of "refutations" from "anomalies" may always be possible, but the problem is whether there are formal rules for delineating these in general. The case studies discussed above provide several examples of nontrivial contradictions in research programmes: acids not containing oxygen for Lavoisier, the problem of specific heats for the kinetic theory of heat, or the precession of Mercury's perihelion for prerelativistic astronomy.[14]

Contradictions in Lakatos are simply moved back within theories of nature, instead of nature itself. Clearly, there is nothing here of the confused "real" contradictions that became central to the naive metaphysics of dialectical materialism, which originated in Hegel's dialectics of nature, or what Alfred Schmidt calls Hegel's "reason submerged in materiality."[15] There are so many bad ideas in the tradition of vulgar Marxism about contradictions—whether in nature, social or historical forces, logic, or some combination of all—that the very category is almost tainted. The idea of a dialectics of nature was also canonical within formerly Communist countries and played a major role in scientific education. Many Western Marxists ignored or renounced these ideas entirely as symptomatic of mechanical and deterministic interpretations of Marx, with Theodor Adorno even remarking that Engels himself was the first vulgar Marxist; in the context of Engels's metaphysical theories, that judgment was largely right.

Recognizing Lakatos's Hegelian debts suggests that this rejection of contradictions in nature, while clearly correct, is due for a reassessment. To see contradictions as now part of theories, language, and just another characteristic of Popper's "objective knowledge" is not novel. What is novel is to give contradictions such a primary role in scientific criticism along with a variety of historicizing techniques.

But Lakatos's reappraisal of contradictions and historicism in science is risky just because of the potential confusions and debilitating precedents in the history of ideas. From our historical perspective we need to return to the problematic status of contradictions in Hegel, and clarify the differences between a dialectics of nature and any notion of real physical, chemical, or biological "contradictions" along with their combinatorial "overcoming" in the natural world. Here too, Hegel developed his own position through his critique of Kant:

> The explanation offered by Kant alleges that the contradiction does not affect the object in its own proper essence, but attaches only to the Reason which seeks to comprehend it. . . . In this way the suggestion was

> broached that the contradiction is occasioned by the subject-matter itself, or by the intrinsic quality of the categories. And to offer the idea that the contradiction introduced into the world of Reason by the categories of Understanding is inevitable and essential was to make one of the most important steps in the progress of Modern Philosophy. But the more important the issue thus raised the more trivial was the solution. Its only motive was an excess of tenderness for the things of the world. The blemish of contradiction, it seems, could not be allowed to mar the essence of the world; but there could be no objection to attach it to the thinking Reason, to the essence of mind.[16]

Hegel lauded Kant for seeing the importance of contradictions for reason/Vernunft, but also wanted to eliminate the regulative barriers Kant used to segregate contradictions from Kantian understanding/Verstand and sensible intuitions. In Kant's tripartite scheme of sensible intuitions, understanding, and reason, while "understanding may be regarded as a faculty which secures the unity of appearances by means of rules, and reason as being the faculty which secures the unity of the understanding under principles . . . reason never applies itself directly to any object, but to understanding."[17] Kant's faculties function as philosophical and cognitive analogues to Newton's methodological rule—his "Rule IV," in fact—in which hypotheses only regulate theories or general propositions, the latter being "derived" or induced from phenomena. For Kant, phenomena must not be contradicted by hypotheses that, in turn, may also not be used in "determining" the phenomena, even as Newton consistently violated this rule in practice.[18] For Kant, reason analogously unifies the manifold of concepts, just as the understanding unifies the sensible manifold. But the relation of "unification" between these Kantian faculties is not transitive: reason is regulative of understanding and never constitutive of experience. Those are Kant's boundaries between dialectical reason and an understanding constitutive of an experiential world, and it is just these boundaries that Hegel wanted to dissolve. Hegel's strategy was to pass the dialectical structure of reason "downward," so to speak, through the faculty of Kantian understanding into phenomenal appearance and hence the world itself, so that, as Hegel said, "external, sensuous motion itself is contradiction's immediate existence."[19]

The analogous deployment of contradictions within science by Lakatos repeats, with Popper, the argument between Kant and Hegel on the role of contradiction, deployed now in the modern discourse of theories and logical relations. Hegel claimed only to be carrying to completion certain ideas about contradiction already introduced by Kant, with Popper, like Kant, introducing a role for explicit contradiction into science and guarding, at crucial points, against "accepting" them.[20] For Lakatos, instead of contradictions being ubiquitous in things, they are ubiquitous in theories, and therefore in need of management by reason. From Kant's perspective, when reason/Vernunft goes beyond its regulative employment of concepts and attempts to apply to phenomenal experience, it engenders pseudorational, transcendent illusions in the form of imaginary convergence points of unconditioned series of causes or ultimate "simples." The transcendental "ideas" that apply to understanding, just as categories that apply to intuitions, as Kant noted, "have an excellent, and indeed indispensably necessary, regulative employment, namely, that of directing the understanding towards a certain goal."[21] Yet the goal does not exist as a possible object of Kantian phenomenal experience; it remains forever a *focus imaginarius,* an ideal point that is never achieved, and that is the point of enforcing boundaries between cognitive faculties. Concepts without intuitions are empty, intuitions without concepts are blind, and reason applied as if it were directed toward true objects of experience—what Kant called the realized *Gegenstand* in contrast to the unreal *Objekt* of reason—produces illusions in the form of an immaterial soul or a fictitious cosmological totality, among Kant's examples of imagined, nonexperiential constructs of reason. These are the boundaries Hegel saw as constraining reason and preventing a conceptualization of contradictions as thoroughly real.

Since Lakatos recaptures Kantian ground in liberalizing the role of contradictions in science, he needs to distance himself from anything like Hegel's historically disastrous account. Lakatos retains some Kantian elements to guard against this prolific mistake in that logical consistency is simply a regulative principle with respect to the world, but not scientific theories. First, the methodology of scientific research programmes applies primarily to theoretical, nonapplied science, and contains little traditional epistemology about how theories describe or represent the world. The body of "most reliable" or "technological" knowledge, Lakatos maintains, is constructed from the body of inconsistent scientific knowledge by truncations, even "mutilations" (*MSE* 219), which provide consistent models of the world in a traditional engineering sense. There is always an additional step as it were, needed to give theories or research programmes a fully realist interpretation. Theoretical knowledge is mediated to the world via applied knowledge, analogous to how Kantian understanding deals with phenomena; the difference, however, is that Lakatos's contradictory reason *is* science as well. When Lakatos remarks that "all theories are born refuted," his substitute for the

Hegelian false is part of the true, these theories are not yet models of the world, and they may require significant changes or "truncations" to engineer consistency. Laws of physics in this sense—Maxwell's electrodynamic laws, for example—"lie," they are "false," but neither are they *not* about the world; they just need to be interpreted in particular contexts to create empirically adequate representations.

For Lakatos, practical and technological knowledge is tied to belief—the belief in ordinary verisimilitude and practical precision—while theoretical reason has little to do with either belief or judgment: "The *practical* rationality involved in constructing the 'body of technological knowledge' is alien to the scientific rationality involved in constructing the 'body of scientific knowledge'" (*MSE* 219n1). Research programmes, as series of propositions with functional components including mathematical and observational theories, and heuristic techniques for moving the programme forward, are as a whole neither true or false; and as temporally variable historical entities, they are not even likely candidates for anyone's simple intentions, such as their beliefs. That is why, as Lakatos observes,

> there is no need to believe (rationally or non-rationally) in the hard core of the program on which one is working. Newton disbelieved his own action-at-a-distance program in its realist interpretation; Maxwell elaborated kinetic theory and Planck quantum theory with actual disbelief. But could not one rationally believe the synchronic cross-sections of a program, "the body of scientific knowledge"? Alas, this body has always been inconsistent. How can one rationally believe an inconsistent set of propositions? (*MSE* 220)

Lakatos replaces the question, "Is theory *T* true?" with "Is theory *T′* an improvement over *T*?" It is precisely the character of those temporal problemshifts that has marks of rationality or not; the "rationality" of belief about whether "theory *T* is true" is an issue for a different type of history. Contradictions are preserved from the world because Lakatos distinguishes theoretical knowledge much like Kant segregates reason. "Consistency," Lakatos writes, "must remain an important regulative principle . . . and inconsistencies (including anomalies) must be seen as problems. The reason is simple. If science aims at truth, it must aim at consistency; if it resigns consistency, it resigns truth" (*FM* 58).

While Lakatos partly returns to Kant to avoid Hegel's realist interpretation of contradiction, he provides no account of experience per se, much less a Kantian one. Unfortunately, no traditional epistemologist will be helped by

Lakatos's answer to the problem of contradictions in nature, as suggested by Ian Hacking when he describes research programmes as a theory of truth without representations, "representation" being a surrogate for any traditional epistemological paradigm.[22] The comparison with Hegel and Kant is nonetheless useful. The reason is that Alfred Schmidt, Jürgen Habermas, and others interpret Marx as balancing, in his depiction of the natural world, a Kantian objectivity that is qualitatively heterogeneous from our understanding of the world, with possibilities for historically specific economies and contingently organized roles for natural resources and human labor—elements of the natural world, if anything is.[23] The gist of this Kantian Marx is that both a historically conditioned understanding of the world and its historically variable reorganizations are substituted for Kant's Verstand. And for Lakatos, too, scientific reason is a historically specific form of producing new theoretical propositions:

> The Scientific Revolution was not marked by a sudden emergence of true or highly verisimilar beliefs replacing false or improbable beliefs. Newtonian science and contemporary relativity theory may well have lower verisimilitude than some of the "wisdom" of Elizabethan times. The Scientific Revolution was marked . . . by the emergence of scientific research programmes and their scientific appraisal. The characteristic of science is not a special set of propositions—whether proven true, highly probable, simple, falsifiable, or worthy of rational belief—but a special way in which one set of propositions—or one research programme—is replaced by another. (*MSE* 222)

Because research programmes are themselves historical, the Kantian "correction" desired by some Hegelian-Marxists is at least roughly present in Lakatos as well. Part of that corrective is that the ability to remake the natural world as part of the social world is "bounded," just as Kant bounded Vernunft. The material limits of existence cannot be completely re-created, and in particular, the contradictions of theoretical science are not real. At the same time, one can specify the limits of understanding the world through historically determinate conditions, instead of Kant's ahistorical conditions for Verstand. The problem of the reality underlying phenomenal appearances, the focus imaginarius of the *Ding an sich,* the "thing in itself," is integrated with history, yet not solved. The "historical conditioning" is partly filled out by the methodology of scientific research programmes, but in response to any urge to project a fundamental "stuff" underneath it all, some underlying permanence in the historical change, Kant is a reminder that *that* is not part of experience. What replaces a dialectics of nature could thus be called "realism with a Kantian-historicist face." While not a prominent element of Lakatos's work, it defines the epistemological-historical direction that he is forced to travel, and that contradictions travel as features of theories, not things.

The Michelson-Morley Experiments

The roles for time and historical phasing in research programme progress culminate in Lakatos's conception of hindsight and his attack on the notion of crucial experiments for research programmes. Lakatos's case for hindsight and against crucial experiments makes use of another piece of great science: the Michelson-Morley experiments and their role, or lack of it, in the emergence of the special theory of relativity.

During the nineteenth century, many physicists assumed the existence of a pervasive and hyperfine "ether" as the medium making possible the transmission of light and, after Maxwell, electromagnetic waves generally. Waves simply were inconceivable without a medium of transmission, and theories of light and ether went hand in hand during the last half of that century. Albert Michelson's, and later Michelson and Edward Morley's, experiments were motivated by the idea that ether either could move along with the earth, or it could be left somewhat behind a moving earth and create a virtual "ether wind." An ether wind would then cause light to travel at slightly different speeds relative to the earth's motion, and Michelson's interferometer was designed to measure such differences with the necessary high precision. If there was an ether wind, a light beam directed "upwind" would take longer to arrive back at its origin, via a set of reflecting mirrors, than a light beam "traversing" the ether. No ether wind would imply that the light beams arrive back at the same time. The moving ether/no ether wind and stationary ether/ether wind theories were associated with, respectively, George Stokes and Augustin Fresnel.

That the speed of light is constant in all reference frames would be one of the remarkable assumptions of the special theory of relativity, so the Michelson-Morley experiments are directly relevant to it. There was once a major debate in the history of science over whether Einstein, at the time of writing his epochal 1905 paper on special relativity, knew of Michelson and Morley's work, or the subsequent analyses made by Lorentz and Fitzgerald. As mentioned already, Popper cites the Michelson-Morley experiments in *The Logic of Scientific Discovery* as a paradigm case of his falsification criterion at work, and other philosophers, historians, or scientists have described the experiments' negative results as "leading to," "compelling," or otherwise "directing" physics toward relativity theory. So it is not a philosophical quibble to assess the experiments' role vis-à-vis theory: to what extent were they seen, or could be seen, as crucial experiments?

Lakatos argues that the Michelson-Morley experiments are a paradigm case of the reinterpretation of observations understood previously through entirely different research programmes, namely, nineteenth-century ether programmes. The consistent result of the Michelson-Morley experiments was that no discernible ether wind was created by the movement of the earth through what was taken to be Newton's absolute space, as assumed in his *Principia*. The methodological and historical question is: when can these null results be taken as evidence against the existence of the ether at all, rather than evidence for or against various ether theories? Lakatos's goal in his reconstruction of the Michelson-Morley experiments is to "explain why crucial experiments are seen to be crucial only decades later" (*FM* 72), thus showing how historical appraisals can distort history by omitting a role for hindsight from science itself, and favoring instead what Lakatos calls "instant rationality."[24]

Unfortunately, Lakatos's account is complicated, even falsified in places, by his research programme reconstruction. Lakatos has the right conclusion: that the role of the entire series of Michelson-Morley, and later Morley-Miller, experiments from 1881 to 1928 in rejecting the ether and "suggesting" special relativity is about zero. Lakatos also apparently came to this conclusion earlier than other historians. We therefore need to disentangle Lakatos's sometimes misleading description from an otherwise accurate methodological diagnosis.

The troublesome historical point for Lakatos's reconstruction is that Michelson was primarily an experimenter interested in the design of measurement apparatus. He did make claims about Stokes's no-wind and Fresnel's ether wind theories, which are reasonably cast as ether programme competitors. But Michelson's motivations were not consistently aligned with settling this debate. His first interferometer experiment in 1881 was intended to measure not the speed of light but the velocity of the earth in absolute space. This idea was Maxwell's, who suggested that a relative ether wind could be used to derive the earth's absolute velocity. Maxwell believed such a measurement to be impossible, recognizing that the relative difference in time would be on the order of 10^{-8}; this is the order of the squared ratio, v^2/c^2, comparing the earth's conjectured velocity to that of light. Michelson's great technical achievement in both 1881 and then 1887 was to be able to detect such minute second-order effects. The 1881 experiment failed to detect an ether wind, and from then on Michelson was caught up in a sequence of confounding interpretations for this and later experiments between Stokes and Fresnel's theories, not to mention the possible explanation using the Lorentz-Fitzgerald contraction factor, $\sqrt{(1 - v^2/c^2)}$, applied to the arms of Michelson's apparatus.

Lakatos's account suggests that Michelson's varying interpretations could perhaps be taken as decisive scientific judgments in the competition between Stokes and Fresnel's theories, while they are relatively insignificant theoretical comments. Michelson is always playing catch-up. He had too many possible explanations to deal with, was not principally a theorist like Lorentz, and was not strongly committed to any perspective except that the ether must exist but, naturally enough, always had some opinion on his strange negative results. Lakatos is right in that the Michelson-Morley experiments never "refuted" the ether until Einstein explained that no ether was needed at all. And many did overinterpret the experiments as "crucial" for relativity theory. Michelson, though, is miscast by Lakatos in the competition among Stokes, Fresnel, and then Lorentz's theories.

Setting aside Michelson's interpretations of his work, Lorentz, one of the last and perhaps greatest ether theorists, did analyze the Michelson-Morley experiments in terms of Stokes and Fresnel's competing theories, much as Lakatos describes: there was an electron-ether research programme. Lorentz's analysis of Michelson's 1881 results showed that the experiment neither proved Stokes's theory, which Lorentz later revealed to be inconsistent, nor refuted Fresnel's. That the results did not corroborate or prove Stokes's theory mattered because Michelson had concluded his 1881 paper with a long quotation from Stokes, and then the comment that "the result of the hypothesis of a [Fresnel] stationary ether is thus shown to be incorrect, and the necessary conclusion follows that the hypothesis is erroneous."[25] In addition to identifying problems with Michelson's apparatus, Lorentz developed a new theory entailing, like Fresnel's, an ether wind, but with lower velocity than the detection limits of Michelson's first interferometer. It was this revised Fresnel theory and combined criticisms of the 1881 experiment that partly motivated the technical changes in Michelson and Morley's famous 1887 experiment.

But there was still more than the theorist Lorentz's criticism influencing Michelson, and only indirectly tied to the Stokes-Fresnel competition as presented by Lakatos. Before the 1887 redesigned experiment came an 1886 noninterferometer experiment for Michelson and Morley: an improved estimate of the speed of light in flowing water. This test implied nothing directly about which ether theory was correct, Stokes's or Fresnel's, but Fresnel used his stationary ether theory to estimate a drag coefficient for the slowing of light by dense optical media. Michelson and Morley confirmed Hippolyte Fizeau's result that a current of water does affect the transmission of light through it: as Fresnel theorized, there was partial drag in a moving medium, although slight, and since that result was based on Fresnel's assumption of an overall stationary ether, the 1886 experiment was also taken as evidence for just that. Far from "refuting" the existence of the ether or questioning it, Fresnel's theory was now taken to be confirmed, or at least supported. As put by the two great Ohio scientists, "The result of this work is therefore that the result announced by Fizeau is essentially correct; and that *the luminiferous ether is entirely unaffected by the motion of the matter which it permeates.*"[26] The 1886 experiment, ignored by Lakatos, does not fit neatly into the competition of Stokes and Fresnel's theories, or ether or electron programmes. Still, it helped reset the stage for the 1887 test, even though governed only indirectly by the question of a stationary or moving ether.

In 1881, Michelson had wanted to estimate the absolute velocity of the earth through space, but in 1887, the objective was to test whether the all-pervasive ether was at rest, at least with respect to the sun of our solar system. Again there was no significant result. That was interpreted by Michelson as refuting both Lorentz and Fresnel's theories, but not ether-wind theories in general. Stokes's theory predicted an ether wind at altitudes high enough to compensate for the earth dragging along the ether close to its surface. Now Michelson really is back between Stokes and Fresnel, much where Lakatos generally puts him. Some velocity differences were observed, yet small enough so that, as Michelson and Morley wrote, "if there be any relative motion between the earth and the luminiferous ether, it must be small; quite small enough entirely to refute Fresnel's explanation of aberration." The next step, then, was not to question the ether but to change the experimental conditions: as the experimenters noted, "At the top of an isolated mountain peak, for instance, the relative motion might be perceptible in an apparatus like that used in these experiments."[27]

Michelson would create an "altitude" version of his experiment in 1897 by constructing a fifty-foot-high elevated pipe frame for taking measurements. Before that, though, Lorentz and Fitzgerald had both proposed the idea of explaining the null results assuming a material contraction of the measurement apparatus in the direction of motion. The contraction hypothesis was not intended by Lorentz to eliminate the ether, however; it was meant to remove the contradiction between Fresnel's theory and Michelson's result. At this point, Lorentz and Michelson were still well within ether programmes, and dueling at their closest for Fresnel and Stokes. The 1897 elevated height test again showed no significant interference fringes. Michelson the experimenter was exasperated: "[It] seems so improbable that one is inclined to return to the hypothesis of Fresnel [!] and to try to reconcile in some other way the negative results."[28] In failing to confirm his favored Stokes's theory using his elevated apparatus, Michelson would not let Lorentz and his contraction have the field, preferring to maintain the contraction hypothesis as one among several explanations. But had *anything* happened by 1897 to suggest the problematic status of the ether? No, nothing, and that is Lakatos's key point. It was also not simply a matter of competing research programmes. It was one research programme, Lorentz's ether-electron programme, against a great engineer and experimenter moving inconsistently and halfheartedly among several theories, none of which questioned the ether as a whole.

After 1887, little more is added to the Stokes-Fresnel choice in terms of experimental design or theoretical interpretation, though a variety of repeat trials were conducted. Michelson took his 1897 results to mean that Stokes's theory came into effect only at even higher altitudes. A high-altitude project was conducted by Dayton Miller in 1921 on Mount Wilson in California, about six thousand feet above sea level. Miller thought that he finally had a positive effect and, in 1924, estimated the solar system's absolute velocity through the ether at approximately two hundred kilometers per second! Michelson carried out a final repeat of his classic measurements in 1928, largely in response to Miller's spectacular claims. Contrary to Lakatos, Michelson was not yet retesting Stokes versus Fresnel. His new apparatus embodied improvements to control for temperature and pressure differences at high altitudes, and again no significant positive result was found. A 1955 analysis of Miller's data by a group led by Robert Shankland finally concluded that uncontrolled temperature variations in the shed on Mount Wilson were responsible for Miller's results.[29] Whatever one believes about the role of the Michelson-Morley experiments before 1905, their subsequent role was only gradually and then confusedly made sense of, and never as clearly "refuting" the existence of the ether.

Again, little of that depends on research programme structure. Nevertheless, the Michelson-Morley and Miller experiments stand as a grand example for the role of hindsight in interpreting experimental results. Lakatos's distinction is to have made hindsight a central feature of scientific methodology, and to have identified an essential aspect of the whole Michelson-Morley story. The interpretation of the experiments often took place within a major ether research programme, with competing theories or subprogrammes on ether wind. These were completely removed from revolutionary ideas about relativity, which could *later* take the Michelson-Morley results as a simple null result corroboration, not a crucial experiment falsifying classical theories of space and time. The moral for Lakatos is that this status

> could not be seen instantly. Even if the experiment was negative, it was not clear, negative exactly to what? Moreover, Michelson in 1881 thought it was also positive: he held that he had refuted Fresnel's but had verified Stokes' theory. Michelson himself and then Fitzgerald and Lorentz explained the result also *positively* within the ether programme. As it is with all experimental results, its negativity for the old programme was established *only later,* by the slow accumulation of ad hoc attempts to account for it within the degenerating old programme and by the gradual establishment of a new progressive victorious programme in which it has become a positive instance. But the possibility of the rehabilitation of some part of the "degenerating" old program could never be rationally excluded. . . . [T]he status of an experiment as "crucial" depends on the status of the theoretical competition in which it is embedded. As the fortunes of the competing camps wax or wane, the interpretation and appraisal of the experiment may change. Our scientific folklore however is impregnated with theories of instant rationality. (*FM* 76, 85)

The spirit of that depiction is exactly correct. It is somewhat misleading to say Michelson "explained" the result, since he was cautious in his support of Stokes, and only weakly contributed toward any research programme. Also, the question addressed in 1881 was the absolute velocity of the earth through space regardless of Michelson's concluding statements about Stokes and Fresnel. Not until after the 1881 and 1886 experiments was the question changed to which ether theory was correct. That still confirms what Lakatos is saying: that you could not "read off" falsifying conclusions from the experiments, the theories "under test" were fluid, and the eventual relevant context for the experiments was nowhere in sight before 1905. Nobody thought these experiments revealed something about the existential status of the ether.

Now is it *that* significant, as Lakatos would have it, that Popper and others misjudged the role of the Michelson-Morley experiments in the genesis of special relativity? Yes, because the misinterpretation was widespread in and outside of science. After 1915, Michelson shifted his attention to projects including tests of the general theory of relativity. Michelson tested the effect of the earth's rotation on the speed of light and found no result. He said that "this result may be explained on the hypothesis of an ether fixed in space, but may also be interpreted as one more confirmation of Einstein's theory of relativity."[30] The *New York Times* turned that into the story line: "Michelson Proves Einstein Theory . . . Ether Drift is Confirmed." After about 1910, the Michelson-Morley experiments were inexorably tied to relativity, either as a refutation of the ether, as providing an inductive basis for the special theory of relativity, or even as "demonstrating" the relative nature of space and time.[31] As the special and general theories of relativity became broadly known, the Michelson-Morley experiments became the starting point for assertions of all kinds, while before they had been of interest only to specialists in optics and electromagnetism. Confusion over what the experiments showed or did not show was rampant, with arguments crossing between physics and the philosophy of relativity. The basic theories and problem settings in which the Michelson-Morley experiments were previously framed, including Stokes's ether-drag theory and Fresnel's stationary ether, were mostly unmentioned.

Well after 1905, the ether was still assumed in explaining optical effects, and waves without a transmitting ether were almost inconceivable. Then, almost suddenly, the experiments were supposed to be ether refuting. Einstein and Michelson were themselves caught up in the rhetoric, even as Einstein made no reference to the Michelson-Morley experiments in his 1905 paper. He famously disposed of any ether as "superfluous," or *überflussig:* "The introduction of a 'luminiferous ether' will prove to be superfluous inasmuch as the view here to be developed will not require an 'absolutely stationary space' provided with special properties, nor assign a velocity-vector to a point of the empty space in which electromagnetic processes take place."[32] That is all he said. The hyperbole associated with the Michelson-Morley experiments after 1905 is inversely proportional to its relevance before 1905 to rejecting the ether. There is also little evidence that Einstein took the Michelson-Morley experiments for granted, as part of the prominent physics of the time. Einstein knew of the Michelson-Morley results via his reading of Lorentz and presumably Lorentz's criticisms of Michelson's Stokes-type interpretations. What is crucial about "what Einstein knew" are Mach's criticism of Newtonian absolute space, Lorentz's electron theory from which Einstein borrowed key elements, the Lorentz-Fitzgerald contraction, and local time coordinates. The experiments play a small role in the emergence of the special theory of relativity, as does the ether per se.

What, then, accounts for the role created for the Michelson-Morley experiments? Relativity was a radical departure for physics, neither easy to understand nor explain, and certainly controversial; Einstein's own heuristic account of his relativity principle compounded the difficulties of saying just what relativity theory entailed. But a "crucial experiment" could be understood equally by a reporter from the *New York Times* or an experienced physicist, and that methodological category helped define an idiom for much of the relativity debate.[33] Lakatos claimed repeatedly that our culture is imbued with tacit assumptions about "instant rationality," and the popular and scientific portrayals of the Michelson-Morley results attest to that tendency.

Contradiction and hindsight thus also complement one another: contradictions or anomalous observations are subject to theoretical reinterpretation that may come before, slightly after, or long after a designed experiment, or not at all, thereby showing "the supreme difficulties in deciding exactly *what* one learns from experience, what it 'proves' and what it 'disproves' " (FM 81). Perhaps it is only after a protracted battle that what was once merely an anomaly *now* receives "the honorific title of refutation, the experiment the honorific title of 'crucial experiment' " (FM 87). Combining Lakatos's treatment of anomalies as faced by a program's positive heuristic with his analysis of hindsight, it is clear why he claims that "*neither the logician's proof of inconsistency nor the experimental scientist's verdict of anomaly can defeat a research program in one blow.* One can be 'wise' only after the event. . . . [Scientific] rationality works much slower than most people tend to think, and, even then, fallibly. Minerva's owl flies at dusk" (HS 113; FM 87).

The examples of Bohr's old quantum theory and the Michelson-Morley experiments provide some history justifying Lakatos's claim. As well, as discussed earlier, the success of theories addressing atomic spectra depend on the phasing of those theories vis-à-vis experimental data. For the Michelson-Morley experiments and their resolution through the special theory of relativity, that temporal phasing was extended to its ultimate form as the reinterpretation of decades of experimentation and anomalous results by striking new scientific ideas. In this way, contradictions were integrated across time with respect to the horizon of hindsight required by current programmes and their resolution of past problems. Lakatos alludes above to the *Philosophy of Right,* where Hegel spoke famously of philosophy as Minerva's owl, which "spreads its wings only with the falling of the dusk." Philosophy, wrote Hegel, "is its own time apprehended in thoughts. . . . Philosophy in any case always comes on the scene too late to give [instruction as to what the world ought to be]."[34] That idea would be criticized by Marx in, for example, the *Theses on Feuerbach:* "The philosophers have only interpreted the world, in various ways; the point is to change it." In other words, while Hegel thought philosophical method was necessarily retrospective and unable to provide substantive forward-looking guidance, Marx took that as a methodological weakness requiring the transformation of philosophy into a normative social theory. Lakatos provides a synthesis of both Hegel and Marx *limited to* the philosophy of science. The owl's retrospective wisdom is placed within science itself, meaning it is scientific, not just philosophical rationality, which has a historical dimension. Science, in this way, is partly its own time apprehended in thoughts.

Even if scientists are unaware of this retrospective rationality, scientific practice contains the kind of backward-looking reason that Hegel thought was usefully supplied by philosophy, at least as implied in Lakatos's reconstructed histories. Marx argued that Hegel's idealism prevented him from correctly understanding the role of history in contemporary politics, and there is a residual idealism in Lakatos, too. This bias is shown when he contends "that relatively few experiments are really important," meaning that, as opposed to the "anti-speculative falsificationists . . . the direction of science is determined primarily by human creative imagination and not by the universe of facts which surrounds us."[35] As mentioned above, Lakatos's account of the Michelson-Morley experiments expresses just this bias as well.

But contrary to Hegel, and more in step with Marx's critical theory, Lakatos is happy to offer, as Hegel said, "instruction as to what the [scientific] world ought to be," since research programmes are appraised as either "progressive" or "degenerating." Indeed, it may be the whole point of a normative reconstruction to assess how well the science of the past was conducted and how well methodologies, particularly Popper's, do at explaining the past. The examples of Young and the early wave theory, phlogiston and oxygen, and nineteenth-century theories of heat provide histories of this type, even as other episodes, like the triumphs of Copernican astronomy and the special theory of relativity, are less usefully described. In any case, as Lakatos observes, a rational reconstruction is not

> just a *selection* of methodologically interpreted facts: it may be, on occasions, their *radically* improved version. . . . One way to indicate discrepancies between history and its rational reconstructions is to relate the internal history in the text, and indicate *in the footnotes* how actual history misbehaved [and how it *ought* to have been] in the light of its rational reconstruction. (*HS* 119–20)[36]

That perspicuous narrative style is the one begun in *Proofs* and then continued in the several historical episodes of "Falsification."

Lakatos thus combines Hegel's intellectualism and Marx's interventionism: they are integrated in Lakatos's historiography, through his value-laden reconstruction and "actual history." This is one important step taken by Marx in his critique of Hegelian hindsight. Marx, as Lukács argues, merely "measured Hegel's philosophy by the [historical] yardstick he himself discovered and systematically elaborated, and he found it wanting" (*H&CC* 17). That is, in his early *Critique of the Philosophy of Right,* Marx simply compared "actual" political and social conditions with that represented in Hegel's rational reconstruction. Lakatos takes the step, analogous to that taken by Marx, of asserting the critical and normative relevance of his own philosophy of science, accompanied by a parallel transformation of Hegel's historical reasoning into science instead of political economy; this shift is also marked by Lakatos's turn *back* to Hegelian ideas in language and not any sort of materialism.[37] Whatever its merits, this dual return to Hegel and Marx is a remarkable conception supporting the autonomy of science *and* the means for its exoteric criticism within a novel and thoroughgoing historicism. The historicism is also more than a posture; it is based on specific histories of science of recent centuries.

For Lakatos it is also, and especially, Popper's philosophy that is measured by history's yardstick. Both the old quantum theory and the Michelson-Morley experiments were pointedly directed by Lakatos at Popper's Kantian-style role for noncontradiction in science, and Popper's insistence on specifying falsifying conditions in advance, not with hindsight. But the clear Hegelian substratum that now is emerging leads to many questions. Just how should Lakatos and Popper be "compared" with respect to the history of science? What standard might be used for that? What does such historical comparison amount to? How is it related to methodological theories of science proper? Also, the obvious Hegelian question is to ask for the historical status of Lakatos's own theory: a historicizing methodology obviously needs to provide some characterization of its own historical being. So we turn next to Lakatos's final face-off between Popper and history, and the explicit role of historiography in Lakatos's philosophical history.

9 | Reason in History

Repeated throughout Lakatos's work is the idea of a difference between historical reconstructions and the actual history on which they are based, and as seen in Lakatos's own reconstructions through his footnote apparatus. In *Proofs,* and to a lesser degree his research programme case studies, the text "is meant to contain a sort of *rationally reconstructed or 'distilled' history. The real history will chime in the footnotes, most of which are to be taken, therefore, as an organic part of the essay*" (*P&R* 5). Real and reconstructed history are related according to Lakatos's "favorite—and by now well-worn—joke that history of science is frequently a caricature of its rational reconstructions; that rational reconstructions are frequently caricatures of actual history; and that some histories of science are caricatures both of actual history and of its rational reconstructions" (*HS* 138).

A superficial interpretation might suggest that one can faithfully distinguish caricatures and "authentic" or uncaricatured history, but that is not forthcoming from Lakatos, nor anyone else for that matter. Historical caricature, so artfully presented in *Proofs,* and even the rational reconstructions of the "Falsification" essay, nonpejoratively stands for the imperfections of historical knowledge and ineluctable dependence of history on some interpretative theory. That is Hegel in a nutshell, too: that consciousness, itself a thoroughly historical product, contains its own contradictions, and actual history is only known imperfectly via an interpreting consciousness. Just as Popper wanted us to acknowledge the fallibility of scientific knowledge, Lakatos wants to do the same for historical knowledge, in which the possible perversions of falsifying caricature can never be escaped: that is Lakatos's skeptical-historical core.[1]

History, Lakatos claims, supports hindsight and contradictions, while their rejection also leads to the falsification of history. Or at least that support is provided by the strange histories appearing in "Falsification," which Lakatos warns are falsifying caricatures, and which he "takes back" as false details presented as historically accurate in the reconstructed text. The technique differs from *Proofs* in that here the histories appear more authentic, even as Lakatos makes apparent their dependence on a specific historiography: it was just this philosophical history that raised the ire of many mainstream historians of science who did not see that Lakatos was exposing their own naïveté about "actual history." A first reconstruction of the chemist William Prout and the idea of isotopes is a somewhat minor historical episode, but its role in Lakatos's essay, it appears, is primarily to introduce Lakatos's rewriting of history and deliberate falsification before the more substantive pieces on Bohr and others. For example, Lakatos almost immediately begins the Prout reconstruction, and a first falsification that Prout "knew very well that anomalies abounded" to his law that atomic weights of pure elements were whole numbers, with a footnote that says, "Alas, all this is rational reconstruction rather than actual history. Prout denied the existence of any anomalies. For instance, he claimed that the atomic weight of chlorine was exactly 36" (*FM* 53n1). Is it appropriate that Lakatos's histories contain falsifications, outright changes from what actually occurred? Perhaps so, if the entire development—the articulation of the series of Poppers into research programmes plus their application to history and against Popper—is intended as a consistent demonstration in medias res of how methodology and history inexorably shape one another. Methodology for Lakatos "is nothing but a rational reconstruction of history, of the growth of knowledge" (*MSE* 178n3).[2] How should that be shown and said?[3]

Lakatos's problem, taken up in "History," is "how the historiography of science should learn from the philosophy of science and vice versa." The main points of the argument are

> that (*a*) philosophy of science provides normative methodologies in terms of which the historian reconstructs "internal history" and thereby provides a rational explanation of the growth of objective knowledge; (*b*) two competing methodologies can be evaluated with the help of (normatively interpreted) history; [and] (*c*) any rational reconstruction of history needs to be supplemented by an empirical (socio-psychological) "external history." (*HS* 102)

These ideas are all addressed by Lakatos's conception that methodologies function as hard cores of *historiographical* research programmes and that these programmes "can be criticized by criticizing the rational historical reconstructions to which they lead" (*HS* 122). The perspective now is that of assessing the application of methodologies to history, and without a vicious infinity of "levels" and metalevels because the research programme "level" just becomes the historiographical metalevel.

Lakatos also continues the criticism of Popper and the normative view on historical reconstruction. Popper asks scientists to state when they will give up their theory; Lakatos finally asks of Popper, "Under what conditions would you give up your demarcation criterion?" (*HS* 123). History as a Marxian yardstick becomes, for the methodology of science, "a test of its rational reconstructions." Or, Lakatos's "test" is a novel and modernized Marxian yardstick for sequences of ideas instead of social institutions. Just as Marx argued that philosophy needed to transform itself into social criticism by comparing the Hegelian political model-ideal with political reality, then so too for the comparison of Popper, or any methodology of science, and history. Lakatos's "evaluation" in (*b*) is his answer to how his and Popper's historiographical programmes should be compared, by jointly comparing the "internal" and "external" histories to which their historiographical programmes give rise: one can compare histories of the Michelson-Morley experiments, Priestly and the phlogiston programme, or other reconstructed histories. While the use of history as a yardstick is classically Marxian, internal and external history mark a neat return to Hegelian historiography unencumbered by any particular philosophy of history, whether Hegelian or Marxist; Marx, indeed, is almost all about the external history of the social and material forces of production, circulation, and distribution, so Lakatos's desire to minimize, though not eliminate, his dependence on external history is sharply anti-Marxist.

Lakatos's historiography depends first on a conception of theory-laden history and the constitutive role of methodologies in interpreting the scientific past. Lakatos gives the most casual introduction to this powerful antipositivist position as a straightforward extension of the theory-ladenness of scientific knowledge. Just as there are no uninterpreted facts for scientists, there are none for historians either. For example,

> the proposition "the Proutian programme was carried through" looks like a "factual" proposition. But there are no "factual" propositions: the phrase only came into ordinary language from dogmatic empiricism. *Scientific "factual" propositions* are theory-laden: the theories involved are "observational theories." *Historiographic "factual" propositions* are also theory-laden: the theories involved are methodological theories. In the decision about the truth-value of the "factual" proposition, "the Proutian program was carried through," two methodological theories

> are involved. First, the theory that the units of scientific appraisal are research programmes; secondly, some *specific* theory of how to judge whether a programme was "in fact" carried through. (*HS* 119n1)

As theory-ladenness is perhaps the single common idea among conflicting diverse accounts of scientific method and practice, it is good to recognize that Lakatos uses it to bridge scientific and historical theories: there are no hard historical facts, and Lakatos is as antifoundational about history as he is about first-order scientific knowledge. At the same time, the theory-ladenness of history is introduced a bit speedily by Lakatos, as he glosses over differences between knowledge of nature and knowledge of past knowledge of nature. A theory-laden observation statement is typically a report of a theory-laden experience with a nonsemantic object, namely nature, while a theory-laden historical statement is a report on a report of a theory-laden experience with a nonsemantic object. Though the distinction cannot be made absolute, we can distinguish between representations of—theory-laden—experience and representations of those representations. There is a significant formal correspondence between theory-laden science and theory-laden history, but not an identity. It is consistent with Lakatos's Hegelianism, however, that he not be overly concerned with that distinction. An "error" of Hegel's, or rather just what his philosophy is, is that he compares forms of knowledge in abstraction from the practices constituting them. Hegel directly asks in the *Phenomenology* how well, for example, art and religion "achieve" the philosophical goal of reflective self-knowledge, even though these practices have many other objectives as well. Hegel is not an antirealist; he just has, as Marx correctly identified in *The German Ideology*, a consistent bias against materialist explanations for anything. Hegel's quasi-logical approach is taken up by Lakatos, and happens to be expressed through Lakatos's quick-and-dirty, albeit largely correct application of scientific theory-ladenness to historical writing.[4]

Another expression of this Hegelian logicism is that Lakatos wants research programmes to be used to describe rationality in different domains of knowledge. Lakatos intends the methodology of scientific research programmes as a general theory of criticism—one that can be used to characterize research programmes in physics, geology, or economics—so it must necessarily be indifferent to the character of the objects discussed by the theories of interest, be they subatomic particles, tectonic plates, or financial markets.[5] Lakatos's creation of historiographical research programmes is another example of this generality with the curious property that it is defined by the self-application of the methodology of scientific research programmes to both the philosophy and history of science themselves. Lakatos abstracts what flows through the logical channels of a research programme from how it flows: if the how of the flow meets research programme criteria—say, even in mathematics—then it is as if mathematics is progressing "quasi-empirically," as Lakatos puts it, like an empirical research programme–driven science.[6] So too, methodologies can be viewed as the hard core of quasi-empirical historiographical programmes. Epistemic conditions across disciplines can be purposely confounded through this approach, and as a general theory of knowledge, the approach commits an epistemological "error." But if one's goal is not epistemological—that is, not to explain how theories represent the world but to make systematic methodological comparisons—then it is acceptable with that limitation. Then a virtue of Lakatos's quasi-empirical application of the theory of research programmes to itself is that irrelevant distinctions between "orders" of knowledge are largely erased. As a consequence of the commonalities between science and historiography, such as theory ladenness, Lakatos's own historical practice qualifies as scientific almost by definition.

Theory ladenness means more than that there is no neutral descriptive history. As discussed earlier, for Lakatos, histories imply normative conceptions of how the scientific past *ought* to have behaved based on the criteria of internal scientific progress. The criteria of progress may be implicit, but they always exist: Popperian falsification, the inductive collection of facts, criteria of simplicity or parsimony, "saving the appearances," Kuhnian normal and revolutionary science, and so on. Such standards, however used, create histories of science, for better or worse, including the methodology of scientific research programmes. Resistance to the idea of theory-laden history is resistance to recognition of this ubiquitous normative content. Of course, observes Lakatos,

> many historians will abhor the idea of any rational reconstruction. They will quote Lord Bolingbroke: "History is philosophy teaching by example." They will say that before philosophizing "we need a lot more examples." But such an inductivist theory of historiography is utopian. *History without some theoretical bias is impossible.* Some historians look for the discovery of hard facts, inductive generalizations, others look for bold theories and crucial negative experiments, yet others for great simplifications, or for progressive and degenerating problemshifts; all of them have some theoretical bias. . . . Long texts have been devoted to the

> problem of whether, and if so, why, the emergence of science was a purely European affair; but such an investigation is bound to remain a piece of confused rambling until one clearly defines "science" according to some normative philosophy of science. One of the most interesting problems of external history is to specify the psychological, and indeed, social conditions which are necessary (but, of course, never sufficient) to make scientific progress possible; but in the very formulation of this "external" problem some methodological theory, some definition of science is bound to enter. History of *science* is a history of events which are selected and interpreted in a normative way. (*HS* 120–21)

Lakatos was charged by Kuhn and others with a conception of history that "is not history at all but philosophy fabricating examples." But for Lakatos, "All histories of science are always philosophies fabricating examples. . . . [E]qually, all physics or any kind of empirical assertion (i.e., theory) is 'philosophy fabricating examples' " (*MSRP* 192). Historical caricature—whether in stereotyped shapes of mathematical or scientific method, or in footnoted-antiphonal rational reconstructions—is Lakatos's distinctive approach to making that fabrication process, as it were, a deliberate topic of study. Historiography is the unifying heuristic of Lakatos's work, and the shared perspective between *Proofs* and the methodology of scientific research programmes is that changes in methodological standards are intrinsically related to historical reflection and historiographical method. The gestalts of mathematical history appearing in *Proofs* as the eighteen dialogue characters, plus the so-called actual history in his parallel footnotes, are then characteristic historical reconstructions, through which normative methodological theories, historical reconstructions, and the alleged actual history of science become inextricably interwoven. The same process is manifest in the "falsified" rational reconstructions of the "Falsification" essay, which intends to show that history never really escapes falsification. This is the import of Lakatos's historiographical theory, more so than whether or not it is used to adjudicate between particular histories of science.

The same role for theoretical categories in reconstructing the past in terms of rational progress, and an antipathy toward naive historiography, is present in Hegel, and with no special dependence on the type of history involved. "We can," Hegel remarked,

> therefore lay it down as our first condition that history must be *apprehended accurately.* But general expressions such as *apprehend* and *accurately* are not without ambiguity. Even the ordinary, run-of-the-

> mill historian who believes and professes that his attitude is entirely receptive, that he is dedicated to the facts, is by no means passive in his thinking; he brings his categories with him, and they influence his vision of the data he has before him. The truth is not to be found on the superficial plane of the senses; for, especially in subjects which claim a scientific status, reason must always remain alert, and conscious deliberation is indispensable. Whoever looks at the world rationally will find that it in turn assumes a rational aspect; the two exist in a reciprocal relationship. (*RH* 29)

Lakatos might have interpolated these words himself into any of his texts! No metaphysical or grand historical scheme is present of the type so often read into Hegel; rather, there is just a straightforward post-Kantian understanding of historical knowledge. Just as for Lakatos history is formed via methodological categories, such as "theories" or "research programmes," Hegel noted that "if, for example, we speak of battles, victories, and the like, these are universal representations which embrace a multitude of deeds, etc. . . . In the sentence: 'The army was victorious,' the whole collection of intermediate steps is fully expressed in a single general representation" (*RH* 211). For both Lakatos and Hegel, their histories are not histories of beliefs or scientific events properly speaking but representations of the past. Lakatos carefully distinguishes events and their representations by remarking, "Unfortunately there is only one single word in most languages to denote history$_1$ (the set of historical events), and history$_2$ (a set of historical propositions). Any history$_2$ is a theory- and value-laden reconstruction of history$_1$" (*HS* 121n1). So too for Hegel: "In our language, the word '*history*' [that is, *Geschichte,* from *geschehen,* to happen] combines both objective and subjective meanings, for it denotes the *historia rerum gestarum* as well as the *res gestae* themselves, the historical narrative and the actual happenings, deeds, and events—which, in the stricter sense, are quite distinct from one another" (*RH* 135). It is good that Lakatos never claimed any originality for his ideas.

Hegel and Lakatos, then, have a shared, almost identical, conception of the theory-ladenness of historiography, which they also tend to see as reconstructing representations and not events; these choices enable them also to reconstruct commonalities across different kinds of knowledge, even as it weakens any implicit epistemology. Moreover, their reconstructions of reason in history share clear limitations in using *only* categories from their philosophical theories, whether it be Hegel's political philosophy or Lakatos's philosophy of science. Such histories are purposefully incomplete. They are to be taken only as "internal" histories making no essential use of contingent historical events to explain the acceptance or rejection of theories, research programmes, or any other normative "unit" of progress. No rational reconstruction is complete for Lakatos without an "external" history consisting of the material, social, political, and ideological causes leading to what occurred: Galileo's persecution by the church, the state control of science in the former Soviet Union, the development of optical technology in northern Europe, the rediscovery of Greek skeptical texts during the Renaissance, or the transmission of Greek science through the Arab world in the Middle Ages. But such historical events, while necessary to explain what occurred, are not used in an internal history as a normative justification. In fact, they may be essential in some cases to explain the irrationalism of what did occur. The internal skeleton may also differ considerably from what actually happened—for example, if constructed strictly following methodological standards as Lakatos liked to do, while the narratives of chapter 7 above make use of research programme categories in a somewhat less exacting manner. As well, the internal history will contain names (for instance, "Bohr [1913]") that only identify documents, not people. The scientist Bohr who created the Bohrian programme exists in the "external" history, and even Bohr's program is not the same as the Bohrian programme, a set of theories—not all of which were taken with equal seriousness by Bohr—to which many contributed.

Hegel has the same complementary notion of sharply circumscribed external history as does Lakatos. There is the same narrowing of subject matter to "eliminate," as Hegel said, "the contingent. Contingency is the same as external necessity, that is, a necessity which originates in causes which are themselves no more than external circumstances" (*RH* 28).[7] The separation in Lakatos's rational reconstructions is just as strong as Hegel's, in that it characterizes what was historically rational: "Indeed, in view of the autonomy of internal (but not of external) history, external history is irrelevant for the understanding of science" (*HS* 102). Internal histories are supposed to be continuous with the theories of the subjects that they are histories of, but again, "internal history is not just a selection of methodologically interpreted facts: it may be, on occasion, their *radically improved* version" (*HS* 119), meaning that the rational reconstruction shows how history *ought* to have "behaved," given a certain conception of progress.

The internal-external distinction becomes operational in the history or philosophy of science, when, for example, an inductivist fails internally to explain the choice of facts for generalization or systematizing. Why certain facts and not others—Kepler had dozens of "laws," to cite one case—are generalized is an external problem for the inductivist. Popper can explain the relevance of certain observational conditions to his internal history as potential falsifiers of refutable theories. But as Popper places unfalsifiable metaphysics—such as Einstein's refusal to accept a probabilistic foundation in quantum physics—outside of science, instead of "within" a programme's hard core, the speculative origins of many theories are relegated to external history and not "part of" a normative reconstruction. Or the proliferation of competing theories is similarly an external problem since it is not motivated by Popper's methodology.

For Lakatos, theory choice may sometimes be explained internally, as dictated by the hard core and positive heuristic, while the choice of corroborating examples or refuting counterexamples is often dictated by the development of the protective belt and status of competing programmes. Priority disputes in science, or who makes a discovery first, for instance, are typically external problems, while for research programmes, the temporal order of predicting facts can be decisive to research programme appraisal. Some priority disputes are internally explicable, or "functional," while others are not, but in all cases an external priority "dysfunction," say the dispute over the discovery of the HIV virus, can be evaluated only against some internal history.[8] External history is still necessary to explain the speed of development of different programs or the repression of a progressive program, as with Trofim Lysenko's pseudoscientific genetics in the Soviet Union. That Lakatos can often explain more science internally does not make his external explanations any easier.

> If a theory is rejected because of a single anomaly, falsificationists need only a weak psychological premise (a sort of Falsificationist Rationality Principle) to explain it as rational rejection. Those who hold that the operative principle is the methodology of scientific research programs have to devise a possibly very sophisticated theory of false awareness to explain—in the same case—the rejection as *rational.* (*MSRP* 191)[9]

And for Lakatos, huge fragments of human knowledge are also relegated to the dustbin of prehistory: the average Elizabethan Englishman or woman, he says, "had no science whatsoever: science was non-existent before the seventeenth century" (*MSE* 222).

The work done by methodologies of science according to Lakatos now becomes clear: they induce normative histories that can be disaggregated into external and internal histories, neither of which is a theory-neutral description of the past. How, then, can Lakatos achieve his goal of claiming progress over Popper, of winning the demarcation game? Lakatos recalls the problems raised for Popper, really Popper$_1$, through his studies of Bohr, the Michelson-Morley experiments, and other historical episodes. Does this history "falsify" Popper's research programme? Yes, the history demonstrates important anomalies or contradictions vis-à-vis Popper's standards for science. And no, since the "naive falsification" of Popper$_1$ is a poor model for science, and we should not then adopt it for assessing philosophies of science. Indeed, all methodologies are simplistically falsified by history, including the methodology of scientific research programmes, just as all theories are born refuted. Scientific historiography makes progress through its sea of anomalies, as does science.

What should then replace Popperian falsification as a historiographical standard? Lakatos's suggestion is his quasi-empirical application of research programmes to themselves, that methodologies of science implicitly define historiographical research programmes. The methodology functions as a hard core characterizing internal histories and, fairly obviously, a positive heuristic by deploying its basic categories, such as "corroborating evidence," "inductive base," "falsification," and so on, along with the relevant logic. Such programmes will not match history perfectly; they will all be falsified somewhere and all will falsify actual history to some extent. But they will also be able to interpret episodes from science's past and help structure external histories necessary to fill out a complete explanation of the past. Two historiographical programmes, Lakatos believes, can be compared through the internal and external histories to which they give rise. Lakatos believes that his historiographical programme explains more of the history of science internally than does Popperian falsification, and with fewer appeals to external history, especially the spurious ones. Lakatos also "predicts" that where scientific progress is taken to have occurred, that progress can be reconstructed via research programme categories and standards. Similarly, in dysfunctional cases such as the Soviet repression of scientific theories, Lakatos "predicts" the need for external history to explain the aberration. He also allows the "cross-questioning" of scientific appraisals. Science thought to be progressive, like Young's wave theory or the molecular-kinetic theory of heat in the 1890s, may be shown not to be so, or vice versa; the case studies of Lakatos and his students indeed revise several appraisals, including those of esteemed scientists, for several important historical episodes.

Such is the normative and critical analogue of the reinterpretation of observations or crucial experiments using alternative scientific theories. In terms of promoting accountable discourse through public rationality on scientific change, Lakatos's approach therefore provides fairly detailed guidance. There must be a starting point for such assessments, and these are taken by Lakatos from the scientific elite, or a received canon of great scientists' work—Robert Boyle, Newton, Michael Faraday, Maxwell, Planck, and Einstein typically, but also Lysenko or Elena Ceausescu, the tyrannical and megalomaniac pseudoscientist wife of Romania's Nicolae Ceausescu before the fall of the Berlin Wall. All these appraisals are open to criticism via research program reconstructions. To use a contemporary example with significant practical implications, one might ask for an exoteric account of the assumed progress of global warming or climate change programmes over their competitors. The answer Lakatos does not want is, "Because the scientists say so." That would be an esoteric elitism that is antithetical to Lakatos's liberal goal of protecting the intellectual autonomy of scientists while holding them accountable for their work. It is exactly Pericles's idea in his funeral oration: "An Athenian citizen does not neglect public affairs when attending to his private business. . . . We consider a man who takes no interest in the state not as harmless, but as useless; and *although only a few may originate a policy, we are all able to judge it.*"[10] In this way, methodologies have the additional function for Lakatos of providing "a code of scientific honesty whose violation is intolerable" (*HS* 103), either for scientists or historians. What is intolerable is not the pragmatic choice of supporting a degenerating programme in science or history—either could "stage a comeback" and it is always a matter of weighing risks against costs—but denying comparative progress or degeneration, and in the demonic and pejorative sense, falsifying history to support that denial.[11] On the other hand, Lakatos's fallibilist approach to scientific history means nothing guarantees against that: no a priori test, like a Popperian crucial experiment, distinguishes false history from the true.

That precarious reality can be kept in mind when considering the reputation Lakatos gained among some philosophers and historians in wanting to use research programme analysis to justify publication rejection, with the elementary particle physics of the 1960s being Lakatos's favorite example of oversupported degenerating research. Here, for instance, is the historian Gerald Holton's take on Lakatos:

> Let us not forget that there is always a strange crew waiting in the wings, anxious to set scientists and scholars aright and to punish what they perceive to be mistakes and heresies. To name just two contenders, there

> are the creationists and, at the other end of the spectrum, people like the late philosopher Imre Lakatos who aimed to set up universal criteria for distinguishing between what he called "progressive" as opposed to "degenerating" research programmes. One of his candidates for the category of degenerating research programmes was contemporary elementary particle physics. He offered his help to funding agencies and editors of scientific journals, so that they might reject "degenerating" work and "lay down statute laws of rational appraisal which can direct a lay jury [for example, Congress] in passing judgment." We can tell here that the fire in the stake is still smoldering, the stake on which all those were burned who ran afoul of an absolutist system.[12]

Holton unfortunately confounds the desirable goal of making scientific criticism exoteric, and the obvious usefulness of research programme categories, with their dogmatic employment. Scientific autonomy is not equivalent to the unconstrained funding and support of any scientific research. The methodology of scientific research programmes is also fairly liberal in that degenerating research programmes may always stage a comeback, so true negative decisions are due to the judgment of decision makers, not a mechanical algorithm.

At the same time, it would be a mistake to think that Lakatos sometimes "just got a little carried away" when he tried to prevent the publication of what he took to be undeserving research in scientific and mathematical topics in which he was not an expert. No: Lakatos's rationalism harbors in caricature the very dangers it is supposed to guard against, and nothing philosophically protects reason from being placed in the service of dogma. In particular, that dogma may be found also in the elite scientists who define the starting point for some of Lakatos's histories. That is, Lakatos *starts* his rewriting of history by taking the judgments of "good science" by recognized scientists, and these normative judgments then may or may not be criticized via methodological and historical analysis. In Hegel, the group playing the role of elites are his famous "world-historical" individuals who, like the proletarian class for Marx and the young Lukács, have a clearer understanding of historical "trends" and may therefore take a vanguard leadership role not possible for others. One of Marxism's antiliberal sides, exemplified by Lenin and the Lukács of *History and Class Consciousness,* was making the Communist Party not just the representative of the vanguard class but an all-powerful elite of the elites. That supposed "dictatorship of the proletariat" never disappeared in Communist states; rather, it simply intensified and encrusted as totalitarianism. Now while Lakatos does not explicitly assert a special "consciousness" for his elites, why else would their judgments be considered special? Clearly, they understand the world differently, otherwise their judgments would not be useful.

These judgments of the scientific elite are also the empirical point of contact for Lakatos's whole historiographical apparatus. The question is: Is the possibility for abusive intellectual power available to Lakatos's elites analogous to either Hegel or Lukács's elites? The difference is that Lakatos provides a detailed means for interrogating, criticizing, and revising their judgments; his is more of a "critical rationalism" than is even found anywhere in Popper. Lakatos's internal-external distinction means that his normative appraisals are not wedded to the elites as people: judgments of good or bad science are affirmed or not via some research programme reconstruction, the latter being the discursive and exoteric representative for what may be only an intuitive understanding. That is all quite hopeful. But the affinities between Lakatos, Hegel, and Marx's elites imply that the inverse of a Popperian "open society"—a closed society in which an ideology of elites is the means for gaining extraordinary, even totalitarian, power—is a close relation of exoteric criticism itself. This is because criticism depends on history, but history is as lacking in certain foundations as science itself. Even the most enlightened historical criticism has no guarantees to protect against falsifications of history justifying the power of elites, but Lakatos's critical apparatus gives some opportunity for protecting against such failures. Perhaps that finesses the truth of Popper's conception of Hegel and Marx's latent totalitarianism in *The Open Society.*

The normative judgments of elites should not be given complete emphasis in Lakatos's histories. It is also possible, as for Hegel or Marx, for the leaders in science to have "false consciousness" about their own historical roles. That is, they have a wrong methodological theory about their own or others' scientific achievements. One remarkable example of such false consciousness in the history of science is represented by Newton. Both Feyerabend and Lakatos see Newton, as Feyerabend says, as a "schizophrenic combination of a conservative" ideology and progressive practice.[13] Even though Newton dogmatically claimed to ground his theories on the stable foundation of experience, his practice, continues Feyerabend, "*is critical* in the sense that it allows for the revision of any part of [science], however fundamental and however close to 'experience.' . . . Thus we have on the one hand the assumption of a stable foundation while we are engaged in an activity that prevents such a foundation from ever coming into existence."[14] The practice developed in spite of its ideology, while the latter did become, for centuries, one ruling conception of scientific method. As Lakatos puts it, "In this sense one may say that, while Newton's method created modern science, Newton's theory of method created [the] modern philosophy of science" (*MSRP* 220). Lakatos and Feyerabend's shared observation is that Newton violated his advice to "avoid hypotheses" throughout his scientific work by just "correcting phenomena" from above, as do all scientists. The combination of method and practice was sometimes successful, sometimes not. As Feyerabend points out,

> It is a marvelous accident, worthy of an equally marvelous explanation that this way of building up the theory of gravitation did not create a miserable patchwork but a coherent system that astounds by its simplicity and its effectiveness in the mastery of concrete phenomena. . . . [L]ight did not submit to this procedure quite as readily. What we obtain here is a system that has been called, and not without justification, "an incoherent and uncertain theory, a theory so full of contradictions and lacunae that one is surprised to see to what extent it could convince the majority of the physicists of the eighteenth century."[15]

The lesson is that while the progress of science is powerfully influenced by scientists' own theories of their practices, these theories may easily be false, or just poor descriptions of what they do. A fish does not understand hydrodynamics, and it should not be assumed that scientists necessarily have the best theories of their own actions. So too, the philosophy of science may be out of phase with scientific practice: in general, Lakatos says, "philosophies of rationality lag behind scientific rationality even today" (*MSRP* 154). In any case, some of "the most interesting historical questions are the interactions of the two—the actual standards and the believed standards."[16] Therefore, ideology, which plays so prominent a role in Marxist theory, and its similar functional relations to practice, is replaced in Lakatos's approach by the methodology of science. There is, though, no "true consciousness" about the methodology of science, no final solution, no group whose judgments about science hold *the* truth protected from criticism. What "constitutes 'false belief' (or 'false consciousness'), depends on the rationality theory of the critic. . . . But no rationality theory can ever succeed in leading to 'true consciousness' " (*HS* 119n3).[17]

10 | A Changing Logic

We move now from the overt comparative role for the methodology of scientific research programmes to the implications of this methodological theory that becomes its own metatheory, or the transformation of the theory of research programmes into a historiographical tool for its own assessment. As we saw, all historiographical research programmes will have their own anomalies or inconsistencies: they will not explain everything, and may lead even to outright falsifications of history, just as Lakatos's own reconstructions sometimes do. Given that, it is even appropriate that Lakatos should ostentatiously "falsify" his rational reconstructions in practice as he is just being a rigorous follower of his own approach. All philosophies of science, in any case, are "born refuted," and anomalies mean that a methodology may be overtaken by a competitor or will generate its own changes—say, through its positive heuristic. Research programmes consist both of changing series of theories and changing interpretations of observations, and that applies to Lakatos's historiographical programme as well: it can and has changed, becoming part of history. The methodology, in this way, posits itself as potentially past, thereby representing a changing logic of scientific discovery.[1] This changing programme is also scientific history because it meets the demarcation criterion it states for what science is, and progressive science to boot, at least relative to Popper. As Lakatos points out, if simple Popperian historiographical falsification is the criterion for evaluating the adequacy of methodologies or philosophies of science, then all methodologies will be falsified by history. In this form, the Pyrrhonian trope of self-application is completely destructive.[2] But when the peritrope is used to turn the methodology of scientific research programmes into a historiographical theory, no inconsistency results through the inclusion of contradictions and potential change. The self-application of a historicist method in that case consistently yields another historicist method, according to which both are transient.

Yet once again, Lakatos's innovation has a clear precedent in the history of ideas. Lakatos's self-application bears comparison to Lukács's controversial application of Marxism to itself in his 1919 essay "The Changing Function of Historical Materialism," published as a chapter of *History and Class Consciousness.*[3] The immediate and obvious consequence of Lukács's formulation, implied in his title, is the transitory and changing function of Marxist theory and ideology. To say the least, this idea was completely opposed to the mechanical and deterministic Marxism prevalent at the time, not to mention stereotypes of Marxist philosophy still common today. Lukács's essay, like Lakatos's "History" piece, begins with the problem of reinterpreting history in light of a new historiographical theory, but here it is the one provided by historical materialism. This theory, writes Lukács, must be turned "into the authentic method for carrying out concrete historical research and for historiography in general" (*H&CC* 223)—analogously, research programme categories are those to be used to write the history of science—while in applying historical materialism to itself, its "substantive truths . . . are truths within a particular social order and system of production" (*H&CC* 228)—that is, Marxist theory is historically specific and variable, too.

Insofar as Marxian method leads to the "self-knowledge of capitalist society," it is self-criticism, or the self-application of theory, that is necessary to continually supply the revolutionary class with "a clear insight into the social preconditions of its own existence" (*H&CC* 229). The Pyrrhonian peritrope, in effect, is here a heuristic of self-knowledge, just as for Lakatos, his self-generated historiographical research programmes are intended to clarify the relations between scientific practice and ideological method. In Lukács, such insights are not available to the bourgeoisie, whose social role depends on a false consciousness of the limitations of its own economic system, while an awareness of these limitations leads either to the system's downfall or outright mendacity—that is, the falsification of history. "The past," writes Lukács, "only becomes transparent when the present can practice self-criticism in an appropriate manner; [quoting Marx:] 'as soon as it is ready for self-criticism to a certain extent, *dynamei* [self-powering] so to speak' " (*H&CC* 225). Lakatos's "self-criticism" similarly attempts to reduce false consciousness in science, vis-à-vis criticism and methodology, and also historicizes his philosophy of science. Rather than describing how to guide theories forward, as in *Proofs,* Lakatos's historiographical research programmes show how entire sets of critical standards may be transformed and made past. For Lukács, his historical perspective exposes the mistake of vulgar Marxism: that "it mistook purely historical categories, moreover categories relevant only to capitalist society [or modern science], for eternally valid ones" (*H&CC* 237). As discussed earlier, Lakatos explicitly associates the theory of research programmes only with modern science; research programme categories also are not "eternally valid." With almost mathematical elegance, when the "base" theory itself includes historical categories, their historicization is the natural consequence of the Pyrrhonian peritrope, whether reinvented as Lakatos's quasi-empirical self-application, or presented as Lukács's changing function for historical materialism.

The most important consequence of Lakatos's changing logic is just that methodological categories and rules are historically variable. His changing logic is intended to dissolve any unhistorical appearance of scientific theories and show that they are subject to history in every respect, including historical decline. History, as a result, does not merely unfold within the terrain mapped out by such methods. It does not resolve itself into the evolution of contents, theories and anomalies, and so forth, while the principles of scientific method remain eternally valid. On the contrary, methodology is precisely the history of these methods, of the changes they undergo *as* methods that bring scientific theories together as research programmes.[4]

It may reasonably be asked, though, Why is it not simply enough to talk about changing logics of discovery or standards of scientific practice, but without the complicated approach? Why isn't it sufficient to consider problems of history and the metaproblems of historical writing, but without the self-application or similar dialectical scheme that creates a theory that is its own metatheory?[5] What motivates the unification of these two problems as one?

The idea originates once again in Hegel's criticism of Kant. While Hegel supported Kant's project to subject metaphysical categories to criticism, Hegel also expected that criteria for the practice of philosophical criticism not escape the critical net.[6] There should be neither an infinite regress of levels and metalevels of discourse nor an implicit appeal to any new, uncriticized standard, such as Kant's "transcendental deduction," whose status as knowledge can be questioned just as Kant challenged, but then of course supported the status of geometry and Newtonian physics. Popper takes the position that methodology has a special quasi-logical status and is not falsifiable as are scientific theories, but from a Hegelian perspective that is just a dogmatic defense. On the other hand, the issue for Hegel was how to evade the skeptical problem of the criterion in general, meaning how to avoid or neutralize an infinite regress of justification. Kant asked in his *Critique of Pure Reason* that the epistemic limits of cognitive faculties be drawn preliminary to their exercise. If this means, as Hegel said, that "the forms of thought must be made an object of investigation," or that thought must itself investigate its own capacity for knowledge, then Hegel agreed with Kant's aim.[7]

Where Hegel diverged from Kant was in a different conception of how such a project should be executed. Hegel wanted the analysis of cognition, and its limitations, to show itself to be a legitimate epistemic act. Whereas Kant demanded, through his apparatus of his categories and the transcendental deduction, as Hegel claimed, "a criticism of the faculty of cognition as preliminary to its exercise," Hegel wanted the critique to be an instance of knowledge as characterized by the critique and to show itself as such.[8] The criticism of philosophical categories should not be "antecedent" to knowledge, but immanent within them. The types of difficulties that Hegel had in mind were, for example, how Kant knew about the relations among the sensible manifold and their intuitions, or the categories of Verstand, or the maxims of Vernunft: for Hegel, the logical-epistemological status of Kant's critical philosophy was unanswered by the *Critique of Pure Reason.* Hegel aimed to close the gap between criticism and actual epistemic claims by uniting the process of philosophical critique with paradigm cases of knowledge. Kant's approach, contended Hegel, concealed "the conception of already knowing before you know—the error of refusing to enter the water until you have learnt to swim."[9] The problem of criticism, and the skeptical problem of the criterion specifically, moves from one place to another; it is not solved.

"True indeed," wrote Hegel, "the forms of thought should be subjected to a scrutiny before they are used: yet what is this scrutiny but *ipso facto* a cognition?" Or stated positively, Hegel's requirement is that one must combine one's metatheory and critical theory proper in the same project:

> So that what we want is to combine in our process of inquiry the action of the forms of thought with a criticism of them. The forms of thought must be studied in their essential nature and complete development: they are at once the object of research and the action of that object. Hence they examine themselves: in their own action they must determine their limits, and point out their defects. This is that action of thought, which will hereafter be specially considered under the name of Dialectic, and regarding which we need only at the outset observe that, instead of being brought to bear upon the categories from without, it is immanent in their own action.[10]

This concept of a complete immanent logic, so characteristic of Hegel, almost to a fault, occurs also in romantic literature, as represented by Friedrich Schlegel's gloss on *Wilhelm Meister* in our epigraph to part 2.[11] The most ancient manifestation of the required logic is expressed, as indicated by Lakatos, through the Pyrrhonian peritrope, the skeptical turning the tables sometimes deployed to trap an opponent in a dialectical debate by showing how they fail to meet their own standard of knowledge in practice. But, as Lakatos calls it, the "Pyrrhonian *machine de guerre*" (*HS* 122) can be positively applied to define a changing logic when knowledge is characterized as fallible and historical. Those conditions are, of course, nowhere to be found in Kant, while quite clearly, in Lakatos's self-application of the methodology of scientific research programmes, "the action of the forms of thought" are combined "with a criticism of them," just as Hegel demanded. This solution to the skeptical problem of the criterion consists largely in recognizing the historical conditions of knowledge. In Lakatos, that means research programmes have a historical character whose status is also that of fallible and changing historiographical tools. The trope of self-application is a means for scientific and historiographical methodology to be unified, and to neutralize, or at least fend off, the problem of the criterion.

Now, does Lakatos's solution then "show" the historic character of modern science? Not exactly. The logic is an exquisite move in the Popperian philosophical demarcation game, yet it is only a self-contained show of its *own* historic character. This philosophical game, which certainly has its own historical character, should not be confused with the historical changes bringing about modern science or the several styles of modern scientific reason in Copernicus, Galileo, or Einstein. Lakatos's changing logic provides a beautiful lesson in dialectical reasoning, even an example of research programme science in action, just as *Proofs is* a continuing series of proofs. In that way, Hegel's goal of immanence is explicitly achieved. But the difference between Lakatos's reflexive historiography and science's history is greater than Lakatos's claim that all histories—including his—are, nonpejoratively, caricatures of history. To paraphrase Kant, more is given through the history of science than can be shown by its rational reconstructions alone: methodologies and rational reconstructions are not exhaustive of science's past, just as Kant wanted to maintain that human understanding and its "sensuous" content remain heterogeneous.

What that ultimately amounts to is impossible to say, but the artificiality of Lakatos's reconstructions, and perhaps any historical reconstructions, prevents them from being more than a changing logic of twentieth-century method, not science itself. To make use of a Kantian distinction noted earlier in chapter 8—that of the real Objekt and the merely logical Gegenstand—what gets historicized through the trope of self-application is a representation of scientific method as a kind of Gegenstand, and again, not science itself. Let's not be fooled by dialectical pyrotechnics: whether scientific methodology has changed significantly over the last two millennia, or even since the time of "the average Elizabethan," is still an empirical question, even as its historical answer demands, as Lakatos correctly insists, *some* theoretical conception of what science is. Change really can be understood without an explicitly changing theory of change. At best, the methodology of scientific research programmes shows why it should or might change, not that it must or even that it has. Thus, again, does the need for a Kantianism with a historicist face become apparent.

That limitation set by real history should not be taken as discounting Lakatos's criticisms of Popper or his numerous historiographical ideas: the theory-ladenness of historical inquiry as well as Lakatos's criticism of an atheoretical and a nonnormative history of science; false awareness by scientists of their own standards of appraisal; prehistory relative to a historical interpretation; elites as sources of historical value judgments; and the critical role of the philosopher in providing historical reconstructions. Another way of considering Lakatos's construction of his changing logic is that he integrates all this by way of a bona fide Hegelian concept, or *Begriff,* in that there is a complete interpenetration of historiographical form with research programme content. Any number of passages from Hegel, "applied" to Lakatos, reveal what he accomplished. For instance, when Hegel observed in his *Science of Logic* that "in every other science the subject matter and the scientific method are distinguished from each other," and "not only the account of scientific method, but even the concept of the science as such belongs to its content, and in fact constitutes its final result," there is an exact fit to research programmes and their historiographical self-application.[12] Hegel's *Wissenschaft* and *wissenschaftlich* are translated as "science" or "scientific," but in Hegel's time they just meant scholarly and rigorous, not scientific in the narrow modern sense. Nonetheless, *both* older and modern senses apply to Lakatos using his concept of science as research programmes.

In Lakatos as in Hegel, there is a complete convolution of form and content, in that Lakatos's method, as Hegel phrased it,

> is not something distinct from its object and content; for it is the inwardness of the content, the dialectic which it possesses within itself, which is the mainspring of its advance. It is clear that no expositions can be accepted as scientifically valid which do not pursue the course of this method and do not conform to its simple rhythm, for this is the course of the subject matter itself.[13]

Similarly, Lakatos's "exposition of what alone can be the true method of philosophical science falls within the treatment of [the] logic itself."[14] Just as Hegel's "*science of logic* has grasped its own Concept [Begriff]," the methodology of science has "grasped" itself via its own theory.[15] The end result, like Hegel's Absolute Idea,

> has itself for subject matter and which, in running itself as subject matter through the totality of its determinations [for instance, via the series of Poppers], develops itself into the whole of its reality, into the system of the science [of logic], and concludes by apprehending this process of comprehending itself, thereby superseding its standing as content and subject matter and cognizing the Concept of the science.[16]

And on and on. I leave it as an exercise for the reader to see that Lakatos's series of Poppers and his self-application provides a quite intelligible interpretation for these strange passages. It is just the simple idea of reflexive criticism, so central to Hegel, that is here Lakatos's key heuristic, as applied to research programme categories. That is the means, starting from a fairly powerful methodological theory, by which the form and content of the methodology of scientific research programmes become one.

The Two Marxes

Today, these endless self-descriptions from Hegel are mostly curiosities for the connoisseur. But they nonetheless demarcate a definitive historical link between Lakatos and nineteenth-century philosophy and social theories. More relevant than the comparisons to Hegel's *Logic* are those that can be made with Marx. Just as Lakatos participates in and criticizes a research programme, so too did Marx claim to extend and critique, all the way through to its inevitable change, classical political economy. Marx's *Capital* is both the further development of the classical research programme and its internal criticism, just as the theory of research programs is an elaboration of Popper's ideas along with their logical and historiographical criticism. As Marx said of his own critique of economics, "It is a presentation of the system and simultaneously through this presentation of it, a criticism of it," thus enacting in social theory Hegel's immanent philosophical criticism.[17]

Only here the categories are not those of Kantian cognition and the philosophical tradition but the social categories of economic organization and their theoretical expression in political economy. Marx's critique becomes an analysis of an economic system via its own categories and demonstrates, or has the goal of demonstrating, that the system as a whole is unstable and leads to the conditions of its own dissolution. Our analogue in Lakatos is that his system—the methodology of scientific research programme plus its self-criticism via historiographical research programmes—also demonstrates the possible conditions for its own supersession. Lakatos's demarcation criterion will be "rejected" when another criterion is proposed that supersedes it according to the metacriterion of historiographical progress. The demarcation criterion and the metacriterion are fundamentally the same, so this says that the methodology of research programmes should be rejected when it is superseded in and through the criticism of its own standards: that again is Lakatos's changing logic. As with Marx, there are not two critiques; rather, there is only one critique with two simultaneous moments.

A key difference with Marx, however, is that his dialectic of capital is not one of ideas in language but of the structure of social relations and their realization as a form of life. The self-generative process that is combined in Lakatos through history and methodology is present in Marx in his analysis of the "general history of the rise of capital," or the "self-determination" or "self-formation" of capital: "We are witness to the process of its becoming through which capital comes into being."[18] Marx's changing logic, then, is one first of labor and capital, and not the logic of their expression in economic theory. Indeed, it is one of Habermas's basic criticisms of Marx that in Marx's social theory there is a one-sided

> reduction of the self-generative act of the human species to labor. Alongside the forces of production in which instrumental action is sedimented, Marx's social theory also incorporates into its approach the institutional framework, the relations of production. It does not eliminate from practice the structure of symbolic interaction and the role of cultural tradition. . . . But this aspect of practice is not made part of the philosophical frame of reference.[19]

Habermas's counterfactual description of what including that aspect could have issued in during the nineteenth century is close to what Lakatos has accomplished:

> On this foundation philosophy's position with regard to science could have been explicitly clarified. Philosophy is preserved in science as critique. A social theory [that is, Marx's] that puts forward the claim to be a self-reflection of the history of the species cannot simply negate philosophy. Rather, the heritage of philosophy issues in the critique of ideology, a mode of thought that determines the method of scientific analysis itself. Outside of critique, however, philosophy retains no rights.[20]

Lakatos's theory of research programmes is just such a critical approach for examining methodology as an "ideology" of science. Lakatos would, I suppose, even disagree with Habermas when he says that there is a great "difference between rigorous empirical science and critique," as it is just criticism, embedded in research programmes, which is the engine for the growth of knowledge.[21] The critique also is exoteric, and intended to rationalize public discourse on science and its progress—an approach that Habermas and many others would endorse at least in principle.

Returning to the role of knowledge, though—specifically economic knowledge in Marx—while surely emphasizing the historical transformations of labor and capital, Marx included economic theory as an integral part of his contemporary world, which would change. Economics was partially ideological for Marx, but much of it was correct as economic science as well, or as correct as could be expected of nineteenth-century social science. How might one reconcile these two views, of bona fide science *and* its historical transience? Can those be reconciled as Lukács's "changing function"? As Lucio Colletti remarks,

> There are two Marxes. On the one hand there is the Marx of theoretical economics who developed and completed as a science the political economy founded by Smith and Ricardo. On the other hand, there is the Marx who is a critic of political economy (not of bourgeois political economy, but of political economy *tout court*), the man who has intertwined (and overturned) the arguments of Smith and Ricardo with a theory of alienation [and history] of which the economists know nothing.[22]

Marx's critique is full court, so to speak. He claimed not only to work within the school of classical political economy but that the principles underlying classical theory—alienated labor as a commodity, surplus labor as the source of surplus values, commodity fetishism, and so on—may be completely superseded, *aufgehoben,* through a change in the form of life making that political economy possible. That these changes occur neither mechanically nor deterministically is one of the central ideas of critical theory, as against the mechanical Marxisms that mostly preceded Lukács and dominated the former Communist states through scientistic ideologies. A less dogmatic and nondeterministic view, enabling "Marxism with a human face," then is left with the problem of characterizing Marx's role as "scientist," but perhaps by explaining away his rhetoric of "science" as a nineteenth-century artifact that may even be to blame for vulgar Marxism and its perverse role in actually existing socialism. There has just been little solace in attempting to jointly conceive of Marx as economic scientist and historicist social theorist, or of reconciling a praxis-oriented conception of social change and action with a so-called scientific method. The latter appears only to move one toward science as Galilean mathematical explanation, not historical consciousness.

But what if science is understood as research programmes along with their changing logic? If one considers Marx's scientific role within nineteenth-century political economy as working within the critical standards of a research programme, then nothing is implied regarding what problems are addressed by the programme: as Lakatos notes, it is not the "what" of flows through the "channels" of a research programme, but the "how." Nothing about a research programme dictates the kind of phenomena explored, whether natural or social, or whether deterministic causal factors are assumed to be at work or not. Now, the important distinction for Lukács and Western Marxism is the constitution of economic facts, meaning that one cannot theorize about, say, commodities without a social life that maintains the institutions of production, exchange, and today, communications and advertising, which make a commodity world possible at all. In contrast is an externally given natural world that is taken to be independent of, among much else, all forms of social being.

The error of dialectical materialism and its variants is exactly to confound social-economic and natural facts; this is one of Lukács's great insights and criticisms for which he would be attacked and forever suspect. But *if* Marx is considered as a Lakatosian scientist—that is, one whose work follows the same methodological criteria found in the natural sciences—the problem of the differing constitution of economic or natural facts needn't even be raised; that just plays no role in the methodology or the demarcation criterion. Marx, or anyone, then is also "free" to theorize on the nature of economic phenomena—as essentially historical, as expressed through specific social relations, as subject to conscious intervention and change, as appearing natural and not a human construction, and so on—in conjunction with his role as a classical economist, as if the laws of political economy were intended as ahistorical descriptions. That combined perspective is scientific, as it were, yet allows for contingent, not deterministic, historical variability. The economic science may also transform itself, just as seen earlier here, through Lakatos's historiographical theory. The Aufhebung of political economy as science and ideology can occur simultaneously with the parallel Aufhebung of the form of life providing political economy with its facts, observation statements, and data, thus showing how a science and its facts may completely dissolve into history. Colletti's "two Marxes"—the son of Hegel, critical and social theorist, and the economic scientist—from this perspective are one. Clearly, then, we finally are ready and required to consider the obvious: the status of Marx's economics and classical political economy as a scientific research programme.

11 | Classical Political Economy as a Research Programme

Marx characterized his relationship to other economists primarily through their consistency with what is called the labor theory of value. He made this theory, taken over primarily from the economist David Ricardo, central to the entire organization of *Capital.* The first volume of *Capital* is based almost entirely on exploring the consequences of Ricardo's approach to the labor theory, and in the third volume, Marx provided what he hoped was an improved version of the labor theory overcoming problems in Ricardo's approach. The labor theory of value may also be thought of as a labor theory of price, in that the theory is intended to explain commodity prices and profits, and their relationship to land and labor as the primary phenomena of early- and mid-nineteenth-century European economies. What, for example, would the effect be of a duty on corn on overall profits in England? Or a change in the wage rate? Should profits go up or down? What are the overall constraints among wages, land rents, interest, profits, and prices, and how do they depend on one another? The labor theory of value is a means for understanding the determinants of prices in market exchange and the distribution of wealth.

In his 1776 *Wealth of Nations,* Adam Smith anticipated the labor theory in many respects. Smith addressed the basic problem of a theory of market prices, or what was the "real measure of exchangeable value," rejecting the mercantilist conception of national wealth as determined by money. Smith instead characterized wealth through the production and consumption of goods, and introduced his notion of natural price. The natural price of a commodity, Smith wrote, "is as it were, the central price, to which the prices of all commodities are continually gravitating," in spite of the constant changes and fluctuations in real market prices.[1] Smith's notion of natural price was intended to help determine, for example, when a change in the underlying value of one or more commodities has taken place, and how that change in value was transformed into prices. Smith made the important distinction between the use value of a commodity, or the practical value it provided in consumption, and the commodity's exchange value, or relative amount by which it sold in the market compared to other commodities. The problem seen by Smith was the need to explain market exchange through some standard measure of commodity value. Such a measure could be held constant against both changing market conditions and conditions of commodity production, and offer the basis for different patterns of market prices.

It was in this context that Smith took labor as a commodity whose value was constant to its owners: "Labor alone, therefore, never varying in its own value, is alone the ultimate and real standard by which the value of all commodities can at all times and places be estimated and compared. It is their real price; money is their nominal price only."[2] Just how Smith's natural price depends on labor and how it influences "nominal" market prices is unclear, even as he introduced his famous conception of the market as the self-regulating practice of multiple entrepreneurs each seeking to maximize only their own interests. Smith additionally proposed three roles for labor as a standard numeraire of value. First is one that Ricardo would take up and extend: that goods are exchanged in proportion to the labor required to produce them. Another suggestion of Smith's, not clearly distinguished from the first, is that goods "command" a certain quantity of labor in exchange, and that prices depend on these amounts commanded. Finally, Smith suggested what is called his "adding up" theory of price: that prices equal the sum of rents, wages, and profits.

With due respect to Smith, it is fair to say that *The Wealth of Nations* contributed many of the founding ideas of classical political economy, but with little theoretical structure or logic through which these ideas could be coordinated. The abstraction of an underlying natural price from variable market prices and the use of labor as a standard measure of value are basic conceptions that both Ricardo and Marx would adopt in their own theories of value, although with greater systematic and arithmetic rigor. At the same time, Smith offered a perspective on economic phenomena and implicit assumptions that would be the basis for the classical school: exchange is conceived in purely functional terms as the transfer of goods and capital; labor, goods, and production inputs are taken as standardized and uniform for the purposes of analysis; and production techniques are assumed as similarly standardized. The market is also taken by Smith, as it would be for Ricardo and Marx, as a completely efficient mechanism for the transfer of capital and the equalization of prices. If profits are higher in agriculture than in manufacturing, Smith assumed that there will be a migration of investment from the latter sector to the former until an industry-wide profit rate is established. Likewise, wages and prices are presumed to gravitate toward uniform values, at least within different economic sectors, such as manufacture or agriculture.

Perhaps most significant in the emerging conceptual scheme for classical political economy was the idea of a productive economy requiring labor, land, and other resources as inputs, and other commodities as outputs. Economic output at a given time then becomes the input for continued reproduction, leading to a continuous cycle of production, exchange and distribution of wealth, and reproduction.[3] Characterizing exchange rates through a common value or numeraire was not new with Smith—the problem appears in Thomas Aquinas, and before Smith, both the British mercantilist and French physiocrat economists characterized value creation through, respectively, exchange and the productivity of land—but Smith's redirection was to take *all* production inputs, including labor, as sources of value.[4] The economy is conceived simultaneously as a physical system transforming labor and nature into commodities over time, and as a social system of prices and exchange.

Both Marx and Ricardo criticized Smith's conceptions of value, and they both made some version of the labor theory of value central to their economics. For example, Ricardo argued that Smith's adding up theory of profit led to the incorrect conclusion that rises in wages or rents would lead profits to rise. Ricardo had a quite different theory of rent, as constituting a portion of profit: as less arable land is brought into cultivation, rents rise and profits decline. In this way, rent limits profit in a way unenvisioned by Smith. At least qualitatively, Ricardo's economics has a theoretical perspective not found in Smith, with the labor theory becoming less a tool for policy analysis and more a technical abstraction, and Marx continued in Ricardo's theoretical footsteps.[5] The problem of explaining changes in the overall profit rate would also be taken up directly by Marx, and using the same labor theory as Ricardo. From Ricardo and Marx's perspective, answering such questions depended on your theory of value, or how you explained that commodities had market prices at all and exchanged in various quantities.

The historical claim, to be outlined below, is that Marx's economic theory is reasonably characterized as a research programme and that he developed his economic theory adopting Ricardo's labor theory of value as the programme's hard core. The programme's goal was to explain the factors governing competitive prices and profit, such as conditions leading to a declining average rate of profit for the economy as a whole, as already examined by Ricardo. The central problem for the labor theory was how to transform labor quantities, as a measure of value, into the prices of commodities observed in the marketplace. Marx developed several approaches to this so-called transformation problem, and it is this series of analytic models that constitutes Marx's elaboration of the labor theory research programme. The programme's heuristic derived from the objective of addressing specific problems, such as a falling profit rate, and canonical techniques, such as the quest for a standard numeraire, both of which occur prominently in Ricardo.

Marx used, throughout the thousands of pages of *Capital,* many ideas developed out of his larger historical and social theories, such as his characterizations of labor and social classes, and their historical development. The labor theory also has natural affinities with Marx's historical theories of class formation and alienated labor. But the labor theory itself is largely synchronic and analytic, with nothing like the broad scope of Marx's general historical, sociological, and philosophical vision. The labor theory and Marx's role as a classical political economist, therefore, are just two components of his larger theory of capital and its social formations. Marx's entire project in *Capital* is, in effect, to substitute his historical and social theory for a naturalistic and ahistorical political economy that neglects the totality of social and historical factors on which contemporary economic events depend.[6] Only the relatively limited project of continuing Ricardo's labor theory as an account of prices and profit may be thought of as a research programme, and certainly not Marx's entire philosophy. Marx's final model, using an analytic methodology as rigorous as Ricardo's, is ultimately unsatisfactory and a failure: the problem of transforming values into prices is not solved by Marx. The economic conjectures or predictions that he derived at best loosely from the labor theory also largely did not come to pass. While some qualified empirical support exists for a falling rate of profit, on the whole, Marx's technical economic predictions were not realized, as is well known.[7]

What can be said positively is that Marx consciously defined a research programme organized around the labor theory of value; argued who was working within or without it, and for what reasons; described the economic phenomena he wanted his theory to explain; and elaborated a series of semiformal structural models, of the kind typical of rigorous mid-nineteenth-century economics. This was his contribution to the labor theory research programme, which was theoretically though not empirically progressive, and continued to degenerate into the twentieth century. In this way, Marx the classical political economist was an economic scientist but his programme failed to progress after his contributions. The labor theory itself did not disappear, though. The transformation problem was taken up again in the 1960s and largely solved by the Cambridge economist and editor of Ricardo's work, Piero Sraffa, in his *Production of Commodities by Means of Commodities.* Sraffa's solution shows that the technical goals set out by Marx—of explaining prices in terms of labor and production factors—can be achieved without a labor theory of value, or really any theory of value, at all. As such, the labor theory has been superseded by this relatively recent competitor.

Ricardo, Marx, and the Labor Theory of Value

The hard core of a programme does not, as Lakatos remarks, "emerge fully armed like Athene from the head of Zeus. It develops slowly, by a long preliminary process of trial and error" (*FM* 48), and the labor theory of value is no exception. Ricardo's 1817 labor theory was motivated by an earlier approach to understanding rent and profits. In an 1815 essay on corn prices, Ricardo had developed a one-sector economic model in which corn is at once the product, eaten by labor for subsistence, and the input for continued production. Ricardo here created a theory on the limits of agricultural profit, arguing that it depended on the difference between the amount of corn produced and that needed just to sustain a labor force, assumed to subsist only on corn. As he put it, the "rate of profits depends on the proportion of production to consumption necessary to such production."[8] The determinants of profit, as it would be for Marx, define a basic problem to be solved. Also occurring in the one-sector corn model is the principal positive heuristic of the classical programme: the use of an invariable measure of commodity value in order to isolate the effects of changes in production—in this case, changes in profitability as they depend on the relative productivity of land of different qualities. In the one-sector corn model, all quantities are expressed as corn, so nothing has to be changed into corn equivalents; there is not yet a transformation problem.

While the model may sound contrived, it had the important consequence of explaining land rent in terms of more fundamental productive relations. Rent was not a new source of revenue in itself, as it was for Smith. Instead, rent was considered as the surplus product of inframarginal land. Rent is a portion of the profit obtained on that land by renters, who from the landowners' position, are just alternative sources of capital and labor for producing corn. Ricardo also maintained that a rising population meant that as cultivation was extended to less productive or marginal land, profits would drop and rents would rise, with a transfer of profits from farmer to landowner. Duties would similarly raise rents and reduce profits by increasing the demand for home-produced corn. For present purposes, it is not the absolute correctness of these arguments that is important but Ricardo's integrated approach to jointly constraining profits and rents, or the raising or lowering of profits, by accounting for changes in corn value through the production process.

When Ricardo needed to defend the role of agriculture as influencing the rate of profit elsewhere in the economy, such as manufacturing, the corn model failed because there was no way to express prices of nonagricultural commodities. Malthus and others quickly criticized Ricardo's one-sector approach for just this reason, and this led to Ricardo's role for labor as his numeraire of value. The search for a numeraire of commodity value, and hence price, is Ricardo's principal positive heuristic, the "sheet anchor" of his system, as he called it. Here is Ricardo's statement of the labor theory in his 1817 *Principles of Political Economy and Taxation:* "The value of a commodity, or the quantity of any other commodity for which it will exchange, depends on the relative quantity of labor which is necessary for its production, and not on the greater or lesser compensation which is paid for that labor."[9] Smith's labor *commanded,* in other words, is to be distinguished from the labor required to produce a commodity, and it is the latter that determines exchange value and hence prices. Labor, asserted Ricardo, "really is the foundation of the exchangeable value of all things, except those which cannot be increased by human industry," and if "we had an invariable standard, by which we could measure the variation in other commodities, we should find that the utmost limit to which they could permanently rise was proportioned to the additional quantity of labor required for their production, and that unless more labor were required for their production, they could not rise in any degree whatever."[10] Again, the point of the labor theory is not just quantitative measurement via a convenient standard nor the "discovery" of value as a metaphysical substratum but to constrain the possible combinations of wages, rents, prices, and profits. Smith equivocated between prices as determined by the labor required for production and labor commanded in the market, and in any case, provided no constraints for price in the adding up theory. Ricardo clearly distinguished labor in production versus exchange and used the labor theory to develop the constraints absent in Smith.[11]

Ricardo's initial statement of the labor theory makes it look as if he simply isolated Smith's "labor required" from "labor commanded." But Ricardo extended Smith's approach, first by identifying the total labor "embodied" in a product—say, the labor required to machine tools for corn production and that needed to clear a field, in addition to the labor required for planting, cultivation, and harvest. When the total labor was accounted for in this way, Ricardo hoped, his additional theory that prices are proportional to embodied labor, or the total direct and indirect labor entailed in commodity production, would enable him to develop a price theory in the spirit of his one-sector corn model. Ricardo's labor theory, then, is characterized by these two elements, which effectively define the hard core to be taken up by Marx: labor embodied is cumulative labor, and prices are proportional to embodied labor.

Ricardo's theory is correct for simple economies involving multiple commodity types and not just a single product such as corn. In modern algebraic notation, one can define many production functions through a matrix of coefficients $\boldsymbol{A}$, and input and output vectors α and β, such that $\boldsymbol{A}\alpha = \beta$, with prices proportional to quantities of embodied labor. Ricardo believed that prices were determined completely by labor and production, while wages and profits were separately constrained: an increase in one led to a decrease in the other. Still, as Ricardo already saw, this approach would not work in general. A necessary condition required for prices to be proportional to embodied labor is that the relative proportions of direct to indirect labor must be constant across different economic sectors such as agriculture and manufacture. That assumption is not realistic, as one may expect early-nineteenth-century agriculture, which was relatively labor intensive, to embody less past labor in its products than would manufacturing, in which considerable past labor is embodied in machinery and other fixed capital. These cumulative labor quantities are also distributed differently in time. Just as future revenues are valued less because of the time value of money or interest, differing time compositions of labor imply different implicit costs of entrepreneurship that then later need to be recouped; yet such differences are not the direct consequence of embodied labor amounts.

Ricardo therefore not only recognized his own labor theory to be false but he also identified the problem with the labor theory as he posed it: the potentially differing relative compositions of labor and capital across multiple economic sectors. Nonetheless, Ricardo was able to make use of a Lakatosian false theory and identify just how the theory appeared to require changes. When Ricardo pinpointed the problematic labor-capital conditions in his *Principles of Political Economy and Taxation*, some readers were led to consider him as finally "not holding" the labor theory.[12] From a research programme perspective that question is somewhat irrelevant. First, there is nothing wrong with building theories on inconsistent research programme foundations, as discussed before. Second, what Ricardo supposedly held is less important than that his model works only for economies with uniform direct-indirect labor ratios. These ratios, renamed the "organic composition of capital" by Marx, are just where Marx extended his approach to prices and value from Ricardo's.

Marx put himself directly in the line of descent from Ricardo in adhering to the labor theory, both in his history of economic theory written in 1861–1863, *Theories of Surplus-Value* and the economic theory proper in *Capital,* the first volume being published in 1867.[13] In volume 1 of *Capital,* Marx developed his economic analysis only going as far as Ricardo's approach in the *Principles,* postponing to volume 3 his own approach for situations in which prices are allowed to diverge from quantities proportional to embodied labor quantities. While Ricardo progressed over Smith in posing and answering structural problems of prices, wages, rents, and profits, a new question asked by Marx was that of the origin of profit. For the economic systems he and Ricardo examined, what ensures a positive rate of profit at all? Marx's famous response was that only labor power is capable of creating additional value in the production process. Hence only labor yields profit, and it is the surplus value contributed by labor, beyond that needed for reproducing the labor force, that is exploited by capitalists. Similar to Ricardo's conception of land rent as a surplus profit, profit itself in Marx is determined by surplus labor. Part of the point of the limited analysis of volume 1 is to explore the simplest economic conditions in which profit can occur—namely, the Ricardian model in which prices *are* proportional to embodied labor. Marx as well as Ricardo understood the limitations of that assumption, but the model was used to lay the foundation for the later analysis in which prices and embodied labor quantities diverge. What was sometimes thought to be a contradiction between the analysis in volumes 1 and 3 is rather the expression of a systematic methodological plan building on Ricardo's labor theory.

Although *Capital* includes hundreds of pages of historical and social-structural accounts of capital formation and distribution, and the historical development of commodities themselves, the underlying simple model of value and prices is still just Ricardo's labor theory. Through this genetic-historical description, Marx thereby provided, among much else, a historical account of economic categories exactly as formulated by the leading political economist of his time. Volume 1 of *Capital is* Ricardo's theory of value as well as a reconstructed historical depiction of that same theory. Marx's dual objective of extending Ricardian theory and showing its historical character could be achieved in no other way. The economic models of volumes 1 and 3 of *Capital* served Marx's goals of contributing to classical political economy and historicizing the discipline itself; the homology to Lakatos's historicization of Popper should by this point also be evident. Marx also knew that

Ricardo's theory of natural price is valid only when each sector of the economy has the same organic composition of capital, and therefore required improvement.[14] In research programme terms, the Ricardian proportional labor theory is the predecessor theory that Marx intended to extend by addressing its central anomaly: the divergence of prices from quantities proportional to embodied labor amounts.

In *Theories of Surplus Value* Marx also discussed Ricardo's critique of Smith; Ricardo's break with Smith's dichotomous "esoteric," or of the labor theory, and "exoteric," or diverging from the labor theory, approaches; and the "new and startling results" produced via Ricardo's criticisms, which "present the whole bourgeois system of economy as subject to *one fundamental law,* and [which] extract the quintessence out of the divergence and diversity of the various phenomena" (TSV 2:169). That "one fundamental law," the labor theory of value, is reasonably characterized as defining the hard core of a Ricardian-Marxian labor theory programme. Marx devoted dozens of pages to explaining Ricardo's isolation of the labor theory as central to the classical programme, and to criticizing Ricardo and others when they deviated from it. It is, indeed, exactly adherence to the labor theory that is the criterion by which Ricardo was considered by Marx to be a good scientist:

> But at last Ricardo steps in and calls to science: Halt! The basis, the starting-point for the physiology of the bourgeois system—for the understanding of its internal organic coherence and life process—is the determination of *value* by *labor-time.* Ricardo starts with this and forces science to get out of the rut, to render an account of the extent to which the other categories—the relations of production and commerce—evolved and described by it, correspond to or contradict this basis, this starting-point; to elucidate how far a science which in fact only reflects and reproduces the manifest forms of the process, and therefore also how far these manifestations themselves, correspond to the basis on which the inner coherence, the actual physiology of bourgeois society rests or the basis which forms its starting-point; and in general, to examine how matters stand with the contradiction between the apparent and the actual movement of the system. This then is Ricardo's great historical significance for science. (TSV 2:166)

Even Ricardo, argued Marx, assumed too much in trying to derive prices and profits as they occur in the market from the labor theory. While value in a capitalist economy should assume only the existence of commodities, Ricardo presumed what should be explained—namely, "wages, capital, profit, the general rate of profit and even . . . the various forms of capital as they arise from the process of circulation. . . . Ricardo introduces not only, *en passant,* the relationship of 'market price' (monetary expression of value) but postulates the whole of capitalist production and his entire conception of the relationship between wages and profit" (TSV 2:168). This criticism is not a polemic; indeed, it continues the type of analysis Ricardo exercised so well against Smith. Ricardo's "scientific necessity," for Marx, consisted in the extent to which profit, rent, interest, and other categories required by a capitalist economy can be developed starting from "the determination of the magnitude of the value of the commodity by labor-time." Ricardo's contrasting "inadequacy" consisted in leading "to erroneous results because it omits some essential links and directly seeks to prove the congruity of the economic categories with one another." In Lakatos's terms, Marx praised Ricardo where Ricardo articulated, from the hard core, the auxiliary theories needed to explain market prices, and criticized Ricardo where these intermediate theories were missing. Marx's conception of the labor theory, therefore, is not that of an isolated empirical hypothesis or theory intended to explain economic phenomena directly but rather the hard core of a programme requiring a protective belt of better and better models. There is nothing, too, in this programme that embodies any particular ethical or political viewpoint, even as it is closely associated with Marx's materialist conception of history. Marx's notion of the labor theory as "science" is that, as he said, the "science consists precisely in working out *how* the law of value operates. So that if one wanted at the very beginning to 'explain' all the phenomena which apparently contradict that law, one would have to give the science *before* the science."[15]

The labor theory is the basis for models of simple economies, just as are contemporary economic theories, and clearly cannot be expected to be perfect. What are today called anomalies or contradictions are here seen as "exceptions" to the labor theory, which Ricardo himself recognized. As noted above, the primary contradiction is that as opposed to Ricardo's own labor theory of value, many commodities did not exchange at prices proportional to their embodied labor quantities, and that was the problem that Marx would take on. Marx's approach would not be, as he said, to "give the science before the science" but it is for just this that numerous economists are criticized in *Theories of Surplus Value,* and in terms that lead interestingly back to Lakatos. The degeneration of the Ricardian school into "vulgar political economy," as Marx dubbed this problemshift, is often due to the ad hoc status of the vulgar theories; that is, in research programme terms, their use of assumptions or theories imported from outside the labor theory. In commenting on Ricardo's exceptions to his law of value, Marx said that

> Torrens takes this discovery of Ricardo as his point of departure, not, however to solve the problem, but to present the 'phenomena' as the law of the phenomena. . . . Here the phenomena manifested in competition [the market] is merely mentioned, registered. Similarly a 'customary rate of profit' is presupposed without explaining how it comes about, or even the feeling that this ought to be explained. . . . Torrens concludes [that] within capitalist production the law of value suddenly changes. That is, that the law of value, which is abstracted from capitalist production, contradicts capitalist phenomena. And what does he put in its place? Absolutely nothing but the crude, thoughtless, verbal expression of the phenomena which has to be explained.[16] (*TSV* 3:71–73)

What is worth noting here is not just the overbearing insistence on the labor theory but what Marx identified as symptomatic of deviations from it. For example, Malthus is castigated for falsifying "his scientific conclusions when they begin to impinge upon class interests. . . . This is his scientific baseness, his sin against science, quite apart from his shameless and mathematical plagiarism" (*TSV* 3:120). Malthus is "considerate towards the ruling classes . . . [that is,] he *falsifies* science for these interests" (ibid.). This dishonesty is not the direct description of a class viewpoint but distortions on the labor theory *through* methodological weakness. Ricardo, in contrast, is honest because he does not shrink from drawing necessary conclusions at variance with class interests: "Ricardo's ruthlessness was not only *scientifically* honest but also a *scientific necessity* from his point of view" (*TSV* 3:118).[17]

Quite straightforwardly, honesty and the falsification of economic history or theory here are indicators of labor theory research programme integrity or its absence for Marx. Methodological rules, writes Lakatos, provide "a *code of scientific honesty* whose violation is intolerable" (*HS* 103), and here Marx can be seen enforcing that normative standard with a vengeance. Indeed, the normative roles for honesty, adherence to the hard core, and falsification of results here play the same part in Marx as in the methodology or research programmes. Here, too, is the original "falsification of history" found later in Marxist dogma, then transferred in *Proofs* to the history of mathematics, and again to the philosophy and history of science in "History."[18] One may certainly disagree with Marx's claims about nineteenth-century economists. But Marx chastised, just as might Lakatos, economists who do not honestly mark divergences from the labor theory when they should. Such is the great tragicomedy of ideas. The reality and rhetoric of falsification would be outdone in the history of Marxism by few other ideas. Its usage in Communist ideology is often an embarrassment or a malicious political act, but it began here with Marx's own correct polemics. What an irony that its source should then turn out to be, in the reasonable sense defined by research programme criteria, "scientific" as proclaimed. A new light is also thrown on Marxian "necessity," as Marx spoke about Ricardo, especially as the methodological similarities with natural science are emphasized.[19] "Necessity" here has a methodological and not a metaphysical connotation; it just signals that an economic explanation coheres with the labor theory rather than "assuming" the phenomena to be explained. In Lakatos's terms, Torrens and other "vulgar" Ricardians continually introduce ad hoc theories of prices and profit relative to the labor programme. None of the connotations of determinism so often attributed to Marx, though with reason, are here.[20]

Marx's role in completing the programme initiated by Ricardo does not make for novel history, as shown by several economic historians: "Marx was the only Ricardian who ever went through with the labor theory of value"; "paradoxically, he can be described as the last, as well as the greatest of the classical economists"; "all doubts that Marx was a great classical economist should have vanished by now"; "for sheer capacity to drive an economic argument to its logical conclusion, Marx was without equal in his own century"; and Marx was, in the end, "Ricardo's only great follower."[21] What is new is to see that Marx's succession can be captured in research programme terms and that his economic theories provide *some* progress over Ricardo.[22] The question then is: What is Marx's "protective belt" of auxiliary hypotheses, which mediates the labor theory as the hard core to market phenomena, and how well does it work? In addition to defining a scientific research programme, does Marx's programme progress as well?

Marx presented in *Capital,* among much else, three model economies using the labor theory, the first two discussed in volume 1 and the third in volume 3. The first model is that of a precapitalist economy that also includes commodities and commodity exchange, and that Marx used to explain the historical and logical conditions required for capital itself. The second and third approaches both include roles for capital: first, one in which profits are proportional to embodied labor, and then, one in which profit is instead proportional to invested capital. The presentation is not descriptive. Marx made use of the same theoretical simplifications and assumptions in Smith and Ricardo's works, such as the rationality of laborers and entrepreneurs in utilizing their resources, the standardization of economic inputs and outputs, and the absence of any intrinsic benefit to labor per se. He also characterized historical stages in the emergence of capitalism to explain, for example, the social conditions needed for commodities to have prices of exchange, then why they exchange at the rates they do, and that profits exist. Such relations among prices and profits are the observations to be explored—Marx's "phenomena"—and also provide potential falsifiers for specific theories.

In Marx's first and second models in volume 1, the labor theory is effectively true by definition in the economies to which they apply. People produce goods that are exchanged either by themselves or through others. In addition, in the second stage of capital, a new class can compel the working class to do more work than "the narrow round of its own life-wants prescribes," and surplus labor makes profit possible.[23] In the third model, prices can diverge from labor quantities, along with the extraction of profit as surplus value. These economies are also assumed to include industrial sectors that can be taken as the basis for an economy sustaining itself through time: agricultural products, sold as commodities sustaining the labor force, and manufactured products, similarly demanded for their own use value as well as to continue the processes of production and consumption. The products have market prices and there is an average rate of profit. How are such quantities determined by labor quantities alone? How are labor quantities transformed into prices and profit?

Marx's Proposed Solution to the Transformation Problem

At the point in *Capital* that Marx was ready to propose his solution to the transformation problem, it is effectively a linear algebra problem in labor quantities, rate of profits, and prices among several industries, or economic sectors, assumed to have varying ratios of direct to indirect labor. Marx did not characterize the transformation problem in that way but he used a numerical example that was intended to be generic, and can be generalized to any number of industries and commodities. Recognizing, as many have for decades, the simple mathematical structure of the transformation problem is startling. Even after the thousands of pages preparing the way for this contribution to the labor theory research programme in volume 3 of *Capital,* very little of Marx's qualitative analyses of labor power, class structure, and commodities plays a direct role in transforming labor values into prices. At bottom, his approach was that of a theoretical economist lacking the benefit of modern mathematical notation and linear algebra.[24]

Constant and variable labor values, c_i and v_i, are taken as given for the *i*th industry producing its commodity, which is not assumed to be an input for the production of other commodities. Although Marx wanted to move beyond Ricardo's prices as being proportional to embodied indirect and direct labor, which here become Marx's *c* and *v* respectively, prices for Marx were "almost" transformed labor quantities, and in that way consistent with Ricardo's approach and that of volume 1. Marx defined market prices as prices of production, and assumed an economy-wide average rate of profit across all industries: as Marx stated, "It is really what Adam Smith calls *natural price,* Ricardo *price of production,* or *cost of production,* and the physiocrats' *prix nécessaire,* because it is in the long run a perquisite of supply, of the reproduction of commodities in every individual sphere."[25] This, then, comes full circle to the many ideas introduced in *The Wealth of Nations* over eighty years earlier.

Marx's assumption that competition for capital across industries leads to a uniform profit rate was also standard since Smith, and mathematically translates into a single profit rate r instead of a profit rate r_i for each industry. The price of production is defined first as the linear sum $p_i = c_i + v_i + r_i\,(c_i + v_i) = (c_i + v_i)\,(1 + r_i)$, with each industry rate of profit, defined in *value* terms, as $r_i = s_i/(c_i + v_i)$, and s_i being the surplus labor over that needed to produce the ith commodity. This is a linear adding up definition of its own, in that production cost is assumed to equal indirect and direct labor quantities, $c_i + v_i$, plus a "markup" for the entrepreneur based on the rate of profit. An economy-wide average rate of profit r is taken as another unknown and is substituted in each industry price, giving $p_i = (c_i + v_i)\,(1 + r)$. To account for the production of a_i units of the ith commodity, each industry is finally represented by an equation $c_i + v_i + r(c_i + v_i) = a_i \cdot p_i$. If there are k industries, then the problem is to solve for the k unknown prices p_i, and the rate of profit r; with labor and commodity quantities taken as given, the problem is to transform them into prices and a rate of profit.

Now the units or dimensions on each side of these equations are different: on the left are labor hours or days, say, and on the right are dollars or English pounds. Hence, some conventional numeraire between these has to be defined. Operationally, the numeraire can be interpreted simply as the money equivalent that would be exchanged for commodities embodying a given equivalent amount of labor. There are only k equations and $k + 1$ unknowns; thus another equation is needed to make the price equations determinate, and that additional equation can be any defining the labor-price standard numeraire. What twist did Marx give to this age-old heuristic in the labor theory programme?

Marx actually overspecified the solution space of prices and the profit rate. He gave two additional conditions that he believed, or assumed, should hold between labor quantities and prices. The first condition is that the sum of all the prices equals the sum of all the labor value quantities, and the second condition is that the sum of the profits equals the sum of the surplus values. As he put it, the "sum of the profits for all the different spheres of production must accordingly be equal to the sum of surplus values, and the sum of prices of production for the total social product must be equal to the sum of its values."[26] This means that Marx was convinced enough of the correctness of the Ricardian labor theory to make labor equal prices and surplus labor equal profit at an overall, economy-wide aggregate level, even if both are false at the industry and individual entrepreneur level due to varying organic compositions of capital, as recognized by Ricardo.

In this way, Marx intended to explain the misleading "appearances" or "phenomena" of the marketplace. Marx tried to transform values into prices by allowing divergences at the commodity level through an economy-wide conservation law, as it were, that normalizes total prices against total labor value *and* total profits against total surplus value. Assuming the equality of these aggregate values and prices is a statement of Marx's belief that labor quantities are redistributed as prices and profits, but in a convoluted form due to the circulation of commodities and capital in the marketplace, and the convergence to a single economy-wide profit rate. Where volume 1 of *Capital* focuses on production, and volume 2 on market circulation, volume 3 is meant to unify both processes using this modified Ricardian model. The solution is an ingenious contribution, in spite of its flaws, to the labor theory research programme. As the economic historian Mark Blaug comments, "The Marxian quest for the appropriate transformation of values into prices is nothing else than the Ricardian hunt for a perfect 'invariable measure of value.'"[27] The solution is to redistribute total surplus value, Marx's source of profit, away from labor-intensive industries with higher organic compositions of capital, to industries with a below-average organic composition of capital. The core of the Ricardian volume 1 analysis is retained underneath the complications of volume 3.

This was truly remarkable theoretical and analytic progress for Marx, especially given the informal use of mathematics in theoretical economics during the early and mid–nineteenth century. Unfortunately, Marx's formulation of the transformation problem, the conclusion of his entire account of the labor theory and the pricing phenomena he and Ricardo wanted to explain, contained a serious and ultimately fatal flaw.

It was first noted by Ladislaus von Bortkiewicz in 1906 that Marx's scheme only transforms values for produced outputs into prices, while the inputs to different economic sectors are still expressed in labor-value terms.[28] An entrepreneur or anyone else purchases products at their prices, not their labor values, and output prices anyway become the input prices for production during the next time step in a cyclical production process. The availability of labor, for example, is observed as its prices, not its values, and these prices should be used to calculate equilibrium commodity prices and the rate of profit. The rate of profit also applies to prices, rather than labor quantities. Can such a revised system of equations in which inputs and outputs are simultaneously transformed into prices be solved? Not if one wants *both* of Marx's conditions to be true together, except in some limited cases. That is the fundamental contradiction, and Marx's research programme progress over Ricardo is therefore almost nil. If inputs are to be transformed as well, one or the other constraint can be selected, but not both. By adding two constraints, Marx overdetermined the price equations and made them insoluble in general. Marx's two constraints are not ad hoc, since they directly reflect his aggregate approach to the labor theory, yet both constraints in general are inconsistent.

Marx's reformulation of the labor research programme is stopped by this serious internal contradiction of its own, notwithstanding its small theoretical progress for the simple economies for which the aggregate model is valid.[29] As such, Marx was an economic scientist in research programme terms. He was, as many have said, truly Ricardo's successor, and he formulated his research programme on a thoroughly consistent basis. Only the research programme could not make more progress than it did. Marx understood the problem of transforming inputs into prices, but he thought that this additional step would not matter significantly.[30] He was wrong.

The story of the labor theory research programme continues, and comes largely to an end, although before reviewing its denouement it is worth looking at one of Marx's economic predictions that is at least consistent with his economic modeling, if not derivable from it. As noted earlier, the question of whether profits would fall or rise under different conditions was addressed by both Smith and Ricardo, the latter arguing against Smith that wages and profits were not independent components of price, but inversely constrained quantities. Marx defined his profit rate as $r = s/(c + v)$, with s = surplus value, c = fixed capital in value terms, and v = variable capital or labor in value terms. By simple algebra, r is also equal to $(s/v)/(1 + c/v)$. Marx believed that, in general, the ratio c/v would rise as a result of technological changes and labor-saving innovations. On its own, a rising c/v would lead to a declining rate of profit. At the same time, Marx expected that increases in productivity would lead to a rising s/v, or rate of surplus value, which would therefore offset the changes in r. By how much? Both ratios will rise, but Marx thought the effect on r of a rise in c/v would win out over the rise in s/v. While the concepts used here are those appearing in Marx's labor theory, in fact, they simply codify an informal theory of technological change that at best serves as a considerable auxiliary hypothesis having little connection to the labor theory per se. As Ronald Meek observes,

> In Marx's time the tendencies he described and analyzed had in fact been revealing themselves on the surface of economic reality—or at any rate were commonly believed to have been doing so—for some considerable time. All Marx really did was to extrapolate these tendencies into the future, on the implied assumption that the relevant economic facts would remain substantially the same and retain the same relative positions on the scale of relevance.[31]

Yet, as Meek says and nearly all have recognized, most of "Marx's 'laws of motion of capitalism' have not revealed themselves on the surface of economic reality."[32]

All that aside, to what extent were at least Marx's extrapolations correct, even if not derived from the labor theory? In his statistical study *The Falling Rate of Profit,* Joseph Gillman concludes that "had our investigations stopped with 1919, the case for Marx's law would have been complete: a rising organic composition of capital; a stabilizing tendency of the rate of surplus value; a falling rate of profit corresponding inversely with a rising organic composition."[33] After 1919, c/v tended to remain constant or fall, while s/v has risen in spurts and r has also risen. Until 1919, then, according to one analysis, "capitalism in manufacturing behaved very nearly as Marx had predicted."[34] That is about the best that can be said for one of Marx's economic predictions, but again, these came from Marx's "ordinary sociopolitical intuition, which did not prove to be much superior to that of his contemporaries."[35] If Marx was an economic scientist, then, even his economic predictions are not empirically relevant to gauging his scientific progress.

There certainly is a historical lesson here, though, regarding Marx's "scientific" status. It is that a naive falsificationist methodology is incorrect for assessing Marx's role in economics. Even as engaged a critic of Marxism as Leszek Kolakowski, in his *Main Currents of Marxism,* said, "As an interpretation of economic phenomena Marx's theory of value does not meet the normal requirements of a scientific hypothesis, especially of falsifiability."[36] That appraisal is a good example of just the phenomenon addressed by Lakatos in his historiographical theory, of how a historical appraisal of "science" will be theory-laden by an implicit methodology. Marx's economic theory had many falsifiable consequences, as did Ricardo's, but it is anyway better understood as a research programme building on the work of Smith and Ricardo, and not a stand-alone single theory. Indeed, when the labor theory and Marx's profit rate are the focus, the situation is actually much worse. If one defines, as Marx did, the rate of profit in *value* terms as $r = s/(c + v)$, but the values are also known to diverge from prices—exactly the problem Marx cared about—then true *money* rate of profit expressed in *price* terms will not be equal to *r*, except as a fluke. While the problem of transforming both inputs and outputs of production into prices can be partially repaired, and even solved using different approaches, Marx's theory of profit is an unrepairable internal inconsistency at the heart of his economics.[37] The rate of profit is completely misconceived as a theoretical measure, notwithstanding problems with inputs and outputs, or an empirically falling rate of profit. And if the theoretical formulation of profit in Marx cannot be used, there is little to say about exploitation, class, and the rest using Marx's actual economic models.

The definition of the profit rate, and the transformation of commodity inputs as well as outputs into prices, are all serious contradictions for the labor theory research programme. Contradictions though, as we saw in the history of the physical sciences, can persist. There has been little additional progress with the labor theory research programme since Marx, so a natural question from a research programme perspective is whether another programme motivated by the same empirical issues can be taken to supersede it.

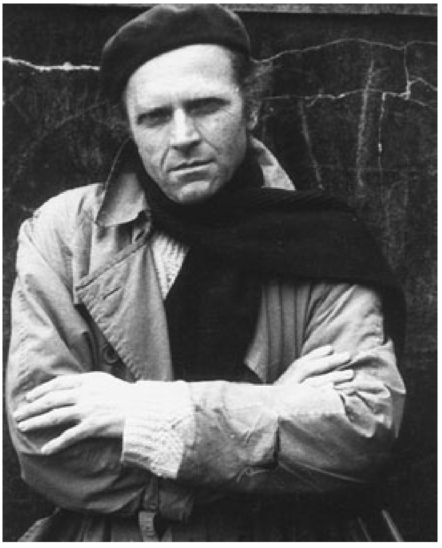
Another way of looking at the labor theory is that it is just one expression of a heuristic to explain market phenomena in terms of production rather than exchange: only the former and not the latter could create commodity value. Competition, similarly, does not create profit from this perspective; instead, it works within a set of constraints for profit-taking opportunities determined by labor and productive relations. So it is not labor per se that motivates the labor theory of value but the goal of limiting one's "givens" to conditions of production as opposed to the market. Given that prices ultimately represent value in whatever sense and labor is essential to transforming raw materials into commodities, some kind of labor-value theory seems like the only way such a programme could proceed. Therefore, some transformation of value into prices seems inevitable. But the labor theory is, in the final analysis, only a means toward the end of explaining prices and profits, and not an independently desired doctrine. What if a competing approach could show that the transformation problem could be solved without a value theory at all?

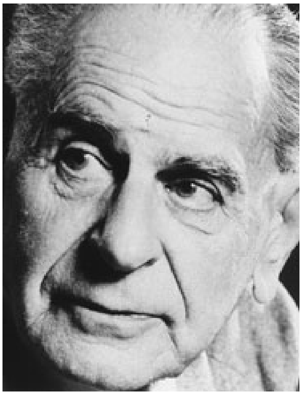
Piero Sraffa achieved that goal with his *Production of Commodities by Means of Commodities* in 1960. Sraffa takes as his starting point only commodities and labor themselves as both inputs and outputs for an iterative production system including a wage rate and rate of profit. In a repeat of Ricardo and Marx's heuristic of identifying an invariant measure, Sraffa defines a standard artificial commodity, as a function of the given commodities, as a numeraire. This measure entails restrictive conditions of its own, and various successors have developed modifications to make the approach more useful.[38] Nonetheless, by returning to the basic Ricardo-Marx problem of value and distribution from the side of production, Sraffa demonstrates that prices and exchange rates are determined using the data that Marx used: the conditions of reproduction of commodities, the real wage, and the distribution of surplus product as a general profit rate. In solving for prices using just the commodity data themselves, value disappears while relations between prices, wages, and profit are explained largely as hoped through the labor theory.

The solution in terms of commodities can be derived also in labor quantities, but really is unnecessary. In Sraffa's approach, a commodity can be represented by the sum of the "dated labor" needed to create it, to create the products creating it, and so on, with past labor "discounted" to the present, analogous to the financial discounting of future revenue streams. As claimed by Maurice Dobb, "What was conceived to be a central problem of classical political economy in Ricardo's day has been solved a century and a half later."[39] It would be fair to characterize Sraffa's work as theoretical progress in classical political economy, though, like many other economic research programmes, awaiting empirical progress as well. It would even appear to be consistent, given the minimal assumptions only of physical commodity data as input, with many other aspects of Marx's general economic theory: alienation, class structure, and so on. For many, including even Nobel laureate Paul Samuelson, the labor theory of value is just redundant, as shown by Sraffa or other nonvalue-based approaches.[40]

To give yet another of the most striking arguments against the labor theory, it is also possible to show that in any economy with a surplus, say corn, the input amount of "corn value embodied" in a unit of corn is less than a unit of corn, as long as profits are positive.[41] If corn is taken as the value numeraire, as it can be, then *corn* instead of labor is the "exploited" commodity. Similarly, any other commodity produced with a surplus can be taken as exploited. This does not show any problem with the idea of surplus production and exploitation; rather, only that it is false to infer that the exploitation of labor surplus explains profit any more than the exploitation of surplus corn, or land, or iron. Even if the labor theory is revived, its use as a theory of exploitation is compromised by the observation that surplus labor is not distinguished economically from any other surplus commodity. Theories of exploitation can still be developed, but independently of a labor or other theory of value. That is largely the contemporary state of progress over the labor theory research programme.

Imre Lakatos.
Clockwise starting upper left, Moscow 1949, 1954, 1961, 1965. All photos courtesy of Éva Pap Davies.





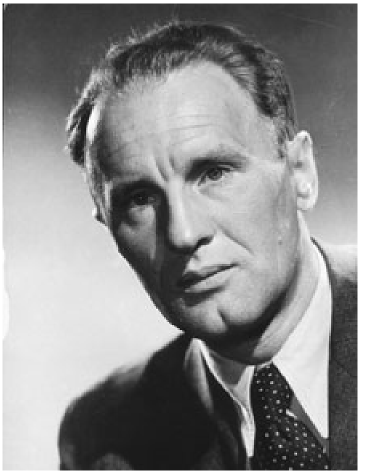
Above: Paul Feyerabend (1924–1994). Lakatos's close friend and most astute critic. Feyerabend's *Against Method* (1974), written as a "long and very personal letter to Imre," had worldwide impact. The book's second half was to have been a rejoinder written by Lakatos, which was made impossible by his death. Courtesy of Grazia Borrini Feyerabend.

Top right: Thomas Kuhn (1922–1996). Author of *The Structure of Scientific Revolutions* (1962), among the most influential English-language philosophy books of the century. Trained as a physicist, Kuhn made empirical history a central feature for the philosophy of science debates of the 1960s and 1970s. A 1965 conference on Kuhn at the London School of Economics became *Criticism and Growth of Knowledge* (1970), edited by Lakatos and Alan Musgrave. Courtesy of MIT Museum.

Bottom right: Karl Popper (1902–1994). Author of the 1934 *The Logic of Scientific Discovery* (*Logik der Forschung*) and Lakatos's mentor at the London School of Economics, Popper came to bitterly resent Lakatos's criticism and transformation of his ideas. An early critic of positivism, Popper was a halfway house to the historicist thinking of Kuhn, Feyerabend, and Lakatos. Popper's *The Open Society and Its Enemies* (1943) cast Plato, Hegel, and Marx as philosophical sources for totalitarianism. Courtesy of London School of Economics.

Georg Lukács (1885–1971). Leading Marxist philosopher of the twentieth century whom Lakatos knew well in Hungary, not always on friendly terms. All of Lakatos's principal Hegelian techniques are prominent in Lukács work. In his *History and Class Consciousness* (1923), Lukács correctly conjectured the existence of a major latent Hegelian architectonic in Marx's work. While wrong on key details, Lukács was largely vindicated by the discovery of Marx's unpublished manuscripts in the 1930s. Lakatos's covert Hegelianism in Anglo-American philosophy has striking and ironic resonances. Lukács is shown here on the right at the 1956 Petőfi Circle debates. At left is journalist and politician Árpád Szakasits. Courtesy of Erich Lessing Archive.

József Révai (1898–1959). Brilliant and dangerous ideological henchman for Rákosi whom Lukács courageously engaged in major public debates. Lakatos's arrest in 1951 and his subsequent years in the Recsk labor camp may have resulted from a plan to denounce Révai as insufficiently Stalinist. Responsible for widespread censorship and agitation-propaganda initiatives before 1956, including Lakatos's subversions at Eötvös College. Courtesy of Hungarian National Museum.



János Kádár (1912–1989). A sometime Stalinist also tortured and imprisoned on Rákosi's orders. Kádár disappeared in the last days of the 1956 revolution, returning as Hungary's Soviet puppet until the fall of the Berlin Wall. Courtesy of Hungarian National Museum.

Mátyás Rákosi (1892–1971). The despised Stalinist leader of Hungary is seen here cynically casting a vote in the 1947 elections. Rákosi considered himself "Comrade Stalin's best Hungarian disciple." Courtesy of Hungarian National Museum.

Imre Nagy (1896–1958). Leader of Hungary during the revolution, then executed after secret imprisonment and trial in June 1958. Lakatos hung a picture of Nagy in his office at the London School of Economics. Shown here addressing the Hungarian Parliament in 1954. Rákosi is at Nagy's right and Ernő Gerő is to his left. Courtesy of Hungarian National Museum.

Imre Nagy. Courtesy of Hungarian National Museum.

Budapest during the 1956 uprising. Courtesy of Erich Lessing Archive.

This page intentionally left blank

III | Magyarország/Hungary

Thus did Naphta astutely go about to turn Herr
Settembrini's paean the wrong way and represent himself
as the incarnation of the cherishing severity of love—
so that it was again impossible to distinguish which side
was in the right, where God stood and where the Devil,
where death and where life.

—THOMAS MANN, *The Magic Mountain*

This page intentionally left blank

12 | Hungary 1956 and the Inverted World

This concluding chapter takes as given the Hegelian content of Lakatos's philosophy of mathematics and science: the bildungsroman form of *Proofs,* the transformation of Popper's philosophy of science into a changing and historicized logic of scientific discovery, Lakatos's unusual historiographical methods, and his critical theory of scientific practice. Lakatos's dual philosophy is put to use here as a means of understanding Hungarian Stalinism between World War II and the failed Hungarian Revolution of 1956, and the emergence of the revolution itself. To do so requires some extended background on the roles of Communist power and ideology in Hungary, and an introduction to several dramatis personae of Hungary 1956. This history will lead back to important themes in Lakatos's English-language work and then to conclusions on his cross-cultural heritage.

Hungary's Stalinist Lebenslüge

Soviet Russia took over Hungary in 1945 following a last destructive siege with the German Nazis who had controlled or occupied Hungary during World War II. Even before the war ended, the Soviets set up a multiparty provisional government that included the Communists led by Mátyás Rákosi, once a junior minister in the failed 1919 Hungarian Communist Revolution, which lasted for less than twenty weeks under the leadership of Béla Kun. Rákosi had returned to Hungary in 1924 to establish a clandestine Communist Party, only to be arrested and jailed with a life sentence. He spent sixteen years in prison, then was extradited to Moscow as part of a 1940 political agreement involving Hungary's borders and just after the most intense phase of Stalin's purges, in which Kun was the principal Hungarian victim. Many other Hungarian Communists, including Lukács, had fled to Moscow during the 1920s and 1930s. Rákosi was their leader in exile and maintained that role on the "Muscovites'" return to Budapest after World War II. From 1945 until 1947, the Party pretense was to combine socialism with features of

Western democracy, and some important reforms were carried out, such as universal suffrage and the redistribution of Hungary's many large feudal-like estates.

By 1949, Hungary had been transformed into a complete Stalinist state. Rákosi achieved Party dictatorship by systematically eliminating the leadership of the three other parties participating in the coalition government, slicing them off one by one through his notorious "salami" tactics. The process was far from purely political. The Soviet army stayed on to provide protection and muscle for the Muscovites, notwithstanding the supposed coalition government. A new Communist-controlled political police, known as the ÁVO, was formed and would become a classic secret police serving the Party. The ÁVO consisted initially of many former members of the Arrow Cross, the Nazi-controlled police before Soviet liberation; included many Jews, even Auschwitz survivors; and was headed by Gabor Péter, the "Hungarian Beria"—Lavrenti Beria being the dreaded head of Stalin's police. Until 1956, the ÁVO would provide Rákosi and the other principal Party leaders—Ernő Gerő, Mihály Farkas, and Jószef Révai, for many years deputy prime minister, defense minister, and minister of education, respectively—with pervasive force to coerce political enemies. Modeling Hungary into the perfect Stalinist state and consolidating Party power was accomplished through bribery, blackmail, public humiliation, intimidation, torture, murder, forced labor, and deportations. Typical Stalinist accusations included complicity with fascism against anyone who worked in the Miklós Horthy government during the interwar years, or even against anyone who had visited Western countries. There was a complete abrogation of justice by the ÁVO, later renamed the ÁVH, which assumed nationwide powers of terror, arrest, and imprisonment directed by Rákosi and his clique. The ÁVO's force was not only hidden in the jails and prisons. When, for example, large companies were nationalized by decree, it was insufficient simply to transfer ownership and reorganize the labor force. Instead, former owners and managers were targeted for degrading hate campaigns, as were smaller tradespeople and shopkeepers compelled to give up their livelihood to the state.

The Party goal was undivided, undisputed, and uncriticizable power throughout Hungary. The late 1940s saw the first major false arrests and fabrications of evidence used to silence, jail, or execute not just political dissenters but Stalinist political rivals and public leaders. Hungary mimicked perfectly the insanity and paranoia of the Soviet purges before the war that had been witnessed firsthand by surviving Hungarian Muscovites. A typically absurd confession was produced after the arrest of the Roman Catholic Cardinal József Mindszenty, in which he "admitted" to being a tool of U.S. espionage, aiming to make Hungary part of a Central European monarchy, and supporting clandestine right-wing movements. The cardinal was jailed with a life sentence, but managed to take refuge in the U.S. embassy from 1956 to 1971. The arrest was then used to facilitate a country-wide purge of all religious leaders. Rabbis and Jewish community leaders were especially persecuted, perhaps to divert attention from the fact that the four Party leaders—Rákosi, Gerő, Farkas, and Révai—were also Jewish, as were many others at the highest ranks of the Party leadership.[1] This irony was just the beginning of the falsity to pervade Hungarian life, and by 1950, about 150,000 people, or 1.5 percent of the population, were incarcerated either in conventional prisons or labor camps. Two thousand are believed to have been executed without trial. Victims of the "Ice Age" purges included many influential and talented professionals, intellectuals, and students, as well as ordinary farmers who did not cooperate with collectivization and central planning efforts.

The greatest of all fabricated show trials, inside or outside the Soviet Union, was that of László Rajk in 1949.[2] Rajk, neither a Muscovite nor a Jew, was a charismatic wartime hero of the underground Hungarian Communists. He was Rákosi's minister of the interior and by no means an innocent Communist. Known as a firebrand, Rajk was capable of insidious tasks when needed. Through Rajk, a witch-hunt and indiscriminate purge began within the Party, with a fabricated Titoist conspiracy supposedly led by Rajk, and motivated by a need to support Russia's break with Marshall Tito, dictator of Yugoslavia, and the expulsion of Yugoslavia from the Soviet bloc. For Rákosi, it was a convenient opportunity also to rid himself of a political rival. Like many other key events in Hungarian history after the war, the Rajk trial was coordinated as a satellite analogue for the Soviet Union's international and internal political struggles, with the ultimate source of power and authority in Moscow. Rajk's trial would combine his stature in the Party and public visibility with an unprecedented, complex plot fabrication based on false confessions extorted from dozens of lesser "accomplices." The trial was supported by numerous in camera inquisitions targeting many of Rajk's colleagues in the ministries, including left-wing socialists who displeased Rákosi for their intellectualism or even expert knowledge of canonical dialectical materialism.

Several steps were involved in the realization of Rajk's conspiracy, meaning the complete fabrication of a historical episode. The ÁVO arrests and torture of supposed accomplices provided confessions and other falsified documents to be used in Rajk's military trial. In Rajk's confession, drafted largely by Révai, he "admitted" to informing for Horthy's police against Communists, betraying comrades in the Spanish Civil War, conspiring with Tito to overthrow the Hungarian government, plotting to murder Rákosi, and intending to restore capitalism. Others were implicated in a plot supposedly orchestrated in Switzerland with John Foster Dulles, the postwar head of U.S. intelligence. Rajk himself had staged rigged trials of his own and knew the consequences of a confession. Refusing to confess, Rajk was visited by his old friend János Kádár, later to become Hungary's turncoat prime minister under Soviet rule after the failed 1956 uprising. Kádár promised Rajk that a confession was purely political support for the anti-Tito campaign, and he would guarantee safety for Rajk and his family. Rajk agreed to the plan and signed the confession. He was hanged on October 15, 1949. The execution was accompanied by a propaganda campaign documenting Rákosi's "sleepless nights" spent uncovering Rajk's conspiracy. Kádár, for his part, fared only slightly better. His conversation with Rajk had been secretly taped by Rákosi for blackmail, and Rákosi decided the next year that his deceitful messenger was better off imprisoned himself. Kádár was also charged with being a foreign agent, and the man who would return as the head of Hungary after the failure of the revolution was sadistically tortured by Defense Minister Farkas's son Vladimir.[3]

Rajk's show trial was the epitome of the combined mendacity and terror pervading Hungarian life. Surveillance and dissemblance were ubiquitous features of everyday life. Apartment block managers were expected to monitor and report on their tenants. Schoolchildren were encouraged to report on their parents, and mail was regularly opened and photostated by the ÁVH. The ÁVH file system included about one million names out of a population of less then ten million, implying a widespread army of voluntary and involuntary informants. Nightfall triggered the fear that an ÁVH automobile could arrive at your home, followed by arrest or deportation at an hour's notice. Economically, the country's central planning model was failing, as it had in the Soviet Union, but ordinary statistics were falsified to hide declines in productivity and food shortages. Collectivization was forced on farmers and peasants while fake Potemkin villages were created for propaganda purposes, including false revenues paid through Budapest ministries. Central planning in Budapest was intense, with rigid production quotas set, say for specific crops, regardless of the suitability of local soils or availability of needed resources. The chain of command went all the way to the Party and Rákosi. Planners having little or no contact with the resources they managed would naturally forget to order materials as trivial as vineyard stakes and trellises.

Planning directives were also enforced by the ÁVH, even in the countryside, with evictions and confiscation of land. "Falsification" then was a learned practice in Hungary, to which even the natural world was not immune to falsification: the weather report for May 1 of any year predicted "no rain" so as not to discourage attendance at the May Day parade.

Stalinist Hungary was a complete *Lebenslüge:* a life of lies in politics, culture, business, industry, agriculture, and everyday life. Word and belief, and word and reality, were systematically polarized through an accumulation of deceit and self-deceit making truth and falsity nearly indistinguishable. Among politicians and the intelligentsia, the Muscovites were most sensitive to the need to dissemble and carefully control one's apparat, as they had learned through years of practice in the Soviet Union. In a bleak inversion of Hegelian logic, Tamás Aczél and Tibor Méray, Hungarian Stalinist writers and later oppositional leaders, wrote in their *Revolt of the Mind* that

> the Muscovite knew that no such thing as permanent truth existed—because no such thing existed in the Soviet Union. He knew that truth has many faces, and the only thing that concerned him was which face was on top just then. He was fully aware that, at all times, truth was what the Secretary-General or the supreme body of the Party held to be truth, and, therefore, it did not particularly bother him that yesterday's truth had changed, by today, into a lie. . . . His explanation was always pat and scientific, always supported by quotations from the Marxist classics. . . . This philosophy he called "dialectics." . . . To the Communists . . . difficulties serve only to be overcome. . . . They knew that everything was true, and the opposite was also true—and it was on this knowledge that they built their lives. (*RM* 11, 13)

The Hungarian lie differed from ordinary mendacity. The deceit of Hungarian Marxist-Stalinism was realized through a Marxist rhetoric of the scientific truths of dialectical materialism, producing a terror state through a parody of Hegelian-Marxist epistemology. For some, the decades of dissembling induced a sort of communist neurosis, meaning the personal inability to distinguish one's true perspective from day-to-day tactics, as falsehood and dissemblance transformed into an intrinsic character trait, grounded in an objective and pervasive paranoia maintained by the police and the culture of surveillance. The falsification of contemporary history was aligned with the dogma of Party infallibility and insulation from internal or external criticism. Rákosi's own past was falsified into a useful hagiography, such as the replacement in history texts of himself for Béla Kun as the leader of the 1919

Hungarian Communist Revolution. Such historical rewriting for the leader's benefit was typical in the Soviet Union. Trotsky, for example, was identified in the *History of the Bolshevik Party* as an imperialist agent already in 1917, the year of the Bolshevik Revolution. The *Great Soviet Encyclopedia* employed editors who visited libraries or schools to cut and paste historical corrections to keep the authoritative reference book "current," and other specialists ensured that individuals appeared or disappeared from photographs or artworks as their fates waxed and waned.[4] Such historiographical methodology had emigrated to Hungary with the Muscovites.

Hungary's political organization complemented the falsification of history and political life. Power was disseminated through an elitist political and social structure emanating from the Party, which was centered on Rákosi as the object of a national personality cult. The paradigm was the classic and then Magyar-perfected Stalinist control model. The idea of personality cult is to associate infallibility and the authority for truth with the cult figure, here primarily Stalin or Rákosi, but also historical figures such as Lenin, Marx, and Engels. The cult figures are a focus at all educational levels, with traditional propaganda techniques mobilized to ensure that cult images are ubiquitous and support the conditioning of behavior and thought. Thousands of Rákosi's portraits hung in offices, classrooms, meeting halls, garages, railway waiting rooms, markets, museums, and editorial offices. The room of a true Communist often had a large Rákosi picture on the wall as well as a smaller desktop photograph. Rákosi called himself by the Homeric epithet "Comrade-Stalin's-best-Hungarian-disciple," and he was surely correct, with the personality cult taking on deep roots comparable to China or Korea. A standard response, for example, when Rákosi's or Stalin's name was first mentioned at a public meeting, or Rákosi's image appeared in a newsreel, was for the audience to break into a rhythmic applause and the recitation "Rákosi! Rákosi! Rákosi!" droning on for minutes.

Psychologically, the combination of imagery and mass behavioral training is a classic propaganda technique, here complemented by Hungary's climate of fear and suspicion. One purpose of the personality cult as a mechanism of power is to turn psychological frailty into a generic submission, so that the cult leader and his cronies are accepted as the bearers of authority and truth. "The essence of the system," insists Polish leader Władysław Gomułka,

> was the creation of an individual hierarchic ladder of cults. . . . In the bloc of Socialist States Stalin stood at the top of this hierarchic ladder. . . . [T]he First Secretaries of the Central Committees of the Parties of the various countries, who sat on the second rung of the ladder of the cult of

> the individual, donned in turn the robes of infallibility and wisdom. . . . The bearer of the cult of the individual was omniscient, knew how to do everything, solved everything, directed everything and decided everything within the sphere of his activity. He was the most intelligent man, regardless of his personal knowledge, capacity, or other personal qualities.[5]

The personality cult is, in this way, the linchpin of a monolithic social epistemology, with knowledge and truth disseminated downward through the cult's political hierarchy: the Party, its administrative channels, and a control network of bureaucracies and decision-making bodies. Elitist, centralized authority is an essential element of a society premised on methodologies of falsehood.

The elitist apparatus held Rákosi as its focus of power and privilege, and emanated outward. The Party as a whole was the most privileged societal "layer"—classes having been allegedly abolished—in a stratified society with politburo and Party members enjoying chauffeurs, elegant villas, vacation homes on Lake Balaton, privileged shopping and schools, plus ÁVH armed guards and barbed wire protection at special compounds. While Rákosi was responsible "up" to Stalin and the Kremlin, the Party had considerable downward power. Less a pyramid than a set of overlapping concentric circles, the Party integrated multiple levels of political power through economic decision making, planning, education, and the arts, so that the circles reinforced one another in the fabrication of truth.[6]

Again, the Rajk trial—in particular, the role of the ÁVH—provides a lesson in the integration of elitist power, control, and the falsification of Hungarian life. By consolidating their capabilities of brute force with technologies of falsification, the secret police became the vanguard of Hungarian Party action during the Ice Age. The ÁVH's role included creating the numerous elements for Rajk's false history: staged conversations, fake confessions, forged documents, reconstructions of phony communications to Yugoslavia, all culminating in the great secret show trial itself. This required much stage setting of the arrests, interrogations, and preparations, through what was called the realization of the trial: the creation of the prefabricated elements and their organization into the conspiratorial narrative and trial scripts. The ÁVH invented its own theory of conspiracy, and then used it to falsify and organize contemporary Hungarian history. As put by Béla Szász, arrested as a colleague of Rajk's and himself a survivor of the trial,

> When the ÁVH investigators were trying to extract some damning statement from a prisoner of long standing or getting him to implicate some-

> one as yet at liberty, they would always demand a *somewhat pointed,* or *politically slanted* deposition. *Pointed* and *politically slanted* belonged to the most frequently used terms in the ÁVH vocabulary, like *contact, essentially* and *realization.*[7]

Through these techniques, documents and depositions were rewritten to exhibit their so-called essential truth to maximum effect. It was through these historiographical methods that the Rajk trial became the most creative and sinister exercise in Hungarian political falsification before 1956. The ÁVH's selective reorganization of synthesized factual detail made the Rajk trial the paradigm of Stalinist Hungary's practical reconstruction of its past. The faked reconstruction of Rajk's past, carried out by the political police as the secret elite within the elite, as well as those with the greatest physical powers of force and terror, epitomized the Party's and Rákosi's dictatorial creation of truth.

In contrast to the relatively uneducated and secret ÁVH, the public elites who scripted everyday Hungarian life were Hungary's many fine and admired writers, poets, journalists, and film and theater directors. Hungary's poets notably have had a tradition of heroic political affiliations and distinguished political works, ranging from Sándor Petőfi and the 1848 war of independence from Austria, to Endre Ady and Attila József in the twentieth century. Hungarian poets and writers were assumed to have pointed and progressive political and social outlooks, and as such, were expected to articulate them through the most beautiful and inspiring resources of the rich Hungarian language. Under Hungarian Stalinism, they also assumed the important propaganda function of disseminating glorified portraits of farm and country life, Rákosi's heroism, the successes of the five-year plan, the historic struggles against fascism, the threats of imperialism, and the promotion of a mechanical Marxism. The writers' intrinsic skills, combined with the traditional esteem in which they were held in Hungary, made them especially well suited to produce Communist ideology in culture, education, and letters. Journalists, too, while not as celebrated as literary figures, supplied the day-to-day expertise needed to describe, for example, the Rajk conspirators and their crimes for the masses when the plot was made public and the stories told of Rákosi's supposedly sleepless nights.

Hungarian writers were not conscripted into Marxism or Communism. Just as in the United States and Western Europe, many Hungarian writers and other intellectuals were attracted to Communism during the interwar years as a response to the emergence of fascism. Hungary, allied with Germany during World War II, was occupied in 1944, and hundreds of thousands of Hungarian Jews were murdered in the last year of the war, many under the direct supervision of Adolph Eichmann. The Hungarian underground Communist Party did not form a significant resistance movement during the war, nor was its membership extensive. But after Germany's defeat, Hungary had no governing model at hand. Its widespread backwardness, including the extensive semifeudal estates worked by a stagnant peasantry, made Marxism attractive as a theory of Hungary's problems and plausible approach for eliminating the tyrannies Hungary had lived under for centuries. Most Hungarian Marxists also did not share the cynicism and dissembling attitudes learned by Muscovites under Stalinism, and were more eclectic and independent, or "left-wing," than their Soviet-trained counterparts in their approaches toward implementing Communism. For many of Hungary's best writers, Marxism and Communism legitimately promised political hope and economic justice.

Thus, many talented novelists, poets, and journalists, including a younger generation that came of age during the war, were actively recruited into the Party as the "transmission belt" of Party propaganda, though they certainly viewed their creative output as describing Hungarian reality. The Hungarian Writers' Association, which would play a central role in the revolution, was formed in 1945 at the initiative of the Communists. Membership was a great honor. All Communist writers joined, and within the association, Communists formed their own elite and politically powerful body. The press and public media generally managed a propaganda apparatus premised on Party infallibility, meaning that all problems had political answers known by the Party. Disseminating those answers through the media was essential to this totalitarian practice and was carried out by the writers who, as Lenin described them, were the "cog and screw" of the Party mechanism.[8]

For this work, the material rewards were the same as for high-ranking Communists. A new and immediately prestigious Kossuth Prize was established for literature, analogous to the Stalin Prize in Russia. Poets became powerful editors, writers were published in deluxe editions unseen in previous years, and magazines and journals were generously funded along with an array of ancillary literary-cum-propaganda posts. All were flattered and enjoyed the status of a new elite exploiting Hungary's traditional ardor for indigenous literature and poetry. Both as reward and requirement, writers were given a coveted position in the Stalinist hierarchy, for brilliantly creating the content flowing through Rákosi's channels of truth.

A principal feature of these authors' work included explaining the transformation of culture brought about since the war by the Party and Soviet model. The assumption was that Russia had been building a socialist state for thirty years and they knew what to do; thus were artful and historically sensitive native styles replaced by mechanical and stereotyped representations of village life, peasantry, the new industrial plan, and the triumph of Communism over fascism. This crude socialist realism was defined negatively by Western bourgeois styles to avoid, plus some heavy-handed political conclusion associated with Party heroics, such as the ongoing "success" of central planning or agricultural policies. The ideological function of journalism and literature was more than cultural window dressing. It was rather one of the most powerful means for amplifying and reinforcing the falsification of Hungarian economic, industrial, and agricultural realities. Most writers had no basis for knowing the actual state of the country outside their Budapest offices. Still, they repeated with belief the information passed on by Party bureaucrats. Poets wrote about labor issues, the struggles of the 1919 Hungarian-Soviet Republic, the new collectivization of land, or the leading role of the working class, usually with sincere motivations, but guided carefully by the instructive criticism of Party managers that was quickly internalized. The traditional aesthetic-political role for Hungarian writing, and the skills of Hungarian writers, had been perfectly co-opted by the Party.

At the top of the cultural hierarchy was József Révai, who with Rákosi, Farkas, and Gerő, was one of the four most powerful politicians in Hungary. Révai was no hack, but a brilliant literary historian and critic who had written fine analyses of Lajos Kossuth, leader of the 1848 revolution, and poet Endre Ady. He was also a servant of Moscow who condemned dozens of writers to years of silence, withdrew "dangerous" books from circulation, and banned innumerable films and plays. Révai was one of the great and pitiful embodied lies of Hungarian culture. Here was a true literary expert and powerful intellectual who would not, for instance, publish a single book by Thomas Mann during his tenure as minister of culture but instead incited major literary debates defining the distorted role for literature in Stalinist Hungary.

One such debate, as it were, occurred during the summer of 1949 as an ideological and cultural component in the realization of Rajk's show trial, and at a time when the Party had attained undivided power. The debate began with an article attacking Lukács written by László Rudas, an official Party philosopher, Muscovite, and founder of the Party academy. After Lukács returned to Hungary from Moscow, he was famous both as a literary critic and philosopher, known for his abiding interest in Hegel. Lukács was admired widely by young intellectuals, not least for his clever resistance to the Party's dreary and stultifying socialist realism.[9] He had high aesthetic standards, yet also said that the social value of literature was to be found not in an author's political affiliations, or literal-minded political readings, or Communist utopian visions, but in the adequacy and honesty with which a novel represented real social dilemmas through fictional means. The writer for Lukács was a partisan in that he or she served the same goals of the Party, but achieved them differently, through literary style and content whose standards were not defined politically. That content was an objective feature of literature, and not a consequence of politically correct authorial identity. Balzac was Lukács's famous example of a conservative writer who nonetheless created a critical realist perspective of the French Revolution and its social aftermath. For at least the younger generation of writers and intellectuals in Budapest to whom Lukács taught literature, Marxism, the history of philosophy, and Hegel after the war, the old Hegelian was venerated as a great teacher and rationalist. Lukács was also seen as tactically and politically shrewd, his knowledge of the arts of self-criticism and survival honed during years of Moscow exile.[10] He was almost always at least slightly at odds with the Party line, and therefore continuously self-criticizing, meaning revising or retracting today what he had written the day or week earlier. Lukács's critical realism and partisan theory of literature were wholly antithetical to the Party's consolidation of power and elimination of criticism.

Rajk's trial, then, gave Révai the opportunity to silence the world's leading Marxist philosopher. Rudas's criticism was to ask whether Lukács, or even the Hungarian Worker's Party, was in a position to provide an interpretation of literary independence and technique differing from Moscow's. Lukács responded to Rudas tactfully, seemingly defending against the attack. But later in the autumn and after Rajk's execution, Révai chose to revive the debate with full force. Lukács's problem, from Révai's perspective, was that he did not truly recognize progress in Soviet literature; Lukács's critical realism disguised latent Hegelian remnants, and thus, Lukács valued good bourgeois poetry over bad socialist poetry. Lukács, it appeared to Révai, supported a tendentious literary style that implied that the poet's partisan role was subversive rather than supportive of the Party. Lukács's relatively common postwar view was that elements of Western democracy would be combined with a socialist economy, and Révai saw this belief as the political root of his incorrect understanding of literature. With Révai as ruler of Hungarian cultural policy making these accusations, it was evident that the esteemed Lukács, then sixty-four, was being used to develop and exploit the literary-ideological side of the Rajk trial.

Lukács would survive the attack, but relevant here is its classic form.

Révai's strategy was to suggest a causal connection between an undesirable political view of the Party with literature, clearly implying that what was needed was Lukács's public admission of error, a recognizable "self-criticism" in which he would criticize and retract his own literary views, admit his failure to grasp the direction and movement of the People's Democracy, acknowledge the value of new socialist literature, and ultimately endorse the ruling regime. Self-criticism was a bizarre ritual of public confession under Stalinism—a stylized speech act realized through publication or some other ostentatious presentation. "These requests for self-criticism," write Aczél and Méray, "are the most characteristic features of Communist Party mechanics. The Party was utterly indifferent as to whether the writers were sincere or not. What the Party wanted was a *statement*" (RM 374).

The practice was all the more eccentric for the Hegelian Lukács, who three decades earlier had argued that Marxism should be critically applied to itself. Under Stalinism, Lukács's clever skeptical trope had been debased into the acknowledgment of power through a ceremonial discursive prostration. When a well-known figure such as Lukács had to exercise self-criticism, he and everyone else knew that his life was on the line, or at least a life in Budapest versus exile or jail. As usual, Lukács managed to weather the public storm, by publishing his self-criticism and withdrawing to his apartment. Such was the "logic" through which Révai silenced one of the major anti-Stalinist voices of the early Ice Age. Though several in Révai's circle apparently still thought Lukács's self-criticism insincere, Lukács was finally left alone. For his part, Lukács reported to some of his students that he "did not exercise self-criticism on the most essential point. . . . In my article, I did not say a single word against Hegel" (RM 79). Lukács was remarkable in these early years of Stalinism for his ability to say the impossible and survive—barely—as Hungary's most erudite oppositional Marxist. He was no Hungarian Stalinist.

Lukács returned for another of Révai's imperious debates, once again illustrating the demeaning and pathetic role of falsification of history in Stalinist Hungary. The protagonist this time was the great Hungarian novelist Tibor Déry, whose book *The Unfinished Sentence* was published after the war; Lukács considered it a major work of twentieth-century literature. Déry had been a Communist in the short-lived 1919 revolution, was jailed under Horthy for translating André Gide's *Return to the U.S.S.R.,* then was awarded the first Kossuth Prize in 1948. Déry, therefore, carried a politically correct history and credentials. His writing was also a fine example of Lukács's critical realism, with Déry describing in detail, for instance, the pillaging of Budapest by Soviet soldiers at the close of the war. Needless to say, such realism was not looked on favorably by the Party. In spite of his fame and earlier acclaim, Déry would be denounced by Soviet censors after critical reviews, and his work banned and pulped.

What had happened, in addition to the realignment of the Communist Party in Hungarian politics, was that Déry had misread the expected ideological role for literature. Révai, a thoroughly compromised intellectual, in spite of his enforcement of propaganda literature could nonetheless rail against "schematism" as the formulaic production of stories according to well-known Soviet models. Déry thought that Révai's disgust with recognizable mediocrity meant an opportunity to create literature attaining high artistic standards and expressing more complex social truths. His mistake, however, was that Révai would always be a cynical enforcer of Party discipline. Déry had begun a four-volume novel called *The Answer,* which chronicled the career of a young working-class man from the 1930s to 1948 under fascism, and ended with the hero, Balint Kope, becoming director of a nationalized factory. When the second volume appeared in 1952, Révai published a merciless review, and a debate on the state of Hungarian literature was announced, it being taken as entirely natural that a government minister could do that. The leader of the pro-Déry faction at the debate was Lukács. But given Lukács's own lack of power, and the evident knowledge that the debate was intended as Déry's intellectual execution, the process was nothing more than a ruthless assault by Révai and his aides on Hungary's leading novelist.

Central to the offensive was Déry's falsification of Hungarian working-class life and history during the 1930s, as represented by Balint Kope's apprenticeship. For example, Révai argued that a historical problem with *The Answer* was that Balint waits until 1945 to join the Hungarian Communist Party. Déry's historically accurate view was that Balint, as a member of the working class, would be unlikely to join an organization that was popular mostly among the petty bourgeois intelligentsia. Révai knew and recognized that Party membership only increased after the end of the war, but Déry's failure to have Balint join earlier nonetheless represented his anti-Party "tendencies." The underground Communist Party of the 1930s and 1940s surely, for Révai, should play a greater normative role in the novel. The expression of personal crises of moral judgment in the book presented another major flaw in that these intimated that problems or circumstances existed that the Party could not solve politically, even as it barely existed historically. As seen by other writers, the upside-down Hegelian

> requirement was obvious. Truth was not that which was true, or what was more, a hundred times true. The writer should not describe a given

> situation or fact or force as it was at the time described. He should portray it according to some "historical truth," taking into account not the momentary situation but some undefinable "future-forming" force. The argument in its simplest form was: Déry falsified by telling the truth; but he would be telling the truth if he falsified. (*RM* 108)

Révai, a nuanced critic, also recognized the undoubted excellence of Déry's work, but the debate was conducted as another ritualized performance: "In those days, Communist criticism had already developed a certain pattern: when criticizing somebody or something, it was obligatory that one also recognized the good points of the person or thing criticized" (*RM* 106). Along with his rewrite of Déry's history, Révai thus provided a kind of immanent criticism, building on Déry's positive ideas, then skewering his victim with accusations of esoteric tendencies or various "isms"—gestalts such as "revisionism" or "cosmopolitanism." The entire absurd rhetoric was directly subservient to Révai's malevolence. The criticisms of Déry's historical falsifications culminated with Révai's effective rewriting of the novel by outlining specific changes to the plot structure and dialogue. To the audience, Révai's arrogant humiliation of Déry was suffused with only slightly disguised threats, the unmistakable message being that Déry should conform as a good Communist. The rewriting of literature and political history reached another apex with the insane production of caricatures of actual Hungarian history executed with élan by an eminent Hungarian literary scholar.

The Road to 1956

So much for several essential elements of Hungarian Stalinism during the Ice Age for our study of Lakatos's philosophy—namely, the pervasive falsification of history and life, the role of elitist power through the personality cult, and the transformation of philosophical tropes into political spectacle. What features of this abbreviated history, then, also drove the country to revolt on October 23, 1956? The relevance to Lakatos's work, as we shall see, continues.

Stalin's death in March 1953 began a slow endgame of Stalinist policies in the Soviet Union and satellite states. Rákosi, Révai, Farkas, and Gerő were called with others to Moscow, where they were criticized for their self-styled "kingdom" and Hungary's abysmal economy. For Moscow, Hungary's security apparatus made conditions ripe for a revolt, and the Hungarian leadership was instructed to change the entire political approach. Rákosi was replaced as prime minister by Imre Nagy, the son of a rural family, a Muscovite though not a leader in exile, and a popular postwar minister of agriculture who had criticized Soviet agrarian policies already in 1949. Révai was demoted and Gerő was moved to the ministry of the interior, but Rákosi retained sufficient power as Party secretary to persist as a rival to Nagy. Back in Hungary, the Central Committee of the Hungarian Worker's Party unambiguously identified Rákosi's faults: "industrialization as end in itself"; "a system of internment that is subject to great arbitrariness"; the "Office of State Security (ÁVH) has been inappropriately led by the party and personally by Comrade Rákosi"; and "instead of a collective leadership, we have direction by one individual, and Comrade Rákosi is largely responsible for the associated cult of personality. . . . Leadership is concentrated in fact in the hands of a foursome: Comrades Rákosi, Gerő, Farkas, and Révai."[11] In July, Nagy gave a speech in the Parliament that astounded all for a similar critical breadth, describing the failures of the five-year plan and Gerő's economic policy, and the country's economic crisis. Nagy's speech was the first sustained public criticism of the last eight years of terror and its consequences, initiating a national "thaw" to last about eighteen months, when Nagy would be asked to step down and himself be expelled from the Party.

For our interests, the thaw in the Ice Age motivated changes among the writers. Some saw the 1953 transformation in the government as reason to recognize and tell truths about Hungary, including the inanities of centralized planning, the phony production statistics, and the continued impoverishment in the villages. This was a start, too, in authors seeing their own responsibility in constructing a completely falsified reality. But while writers and journalists began to recognize that the realities of Hungarian life were "falsifying" their representations of it, conversion was neither uniform nor especially rapid. Many stuck with Rákosi and the hard-liners that Nagy and his followers were the ones "falsifying reality." How much of that was due to actual belief versus tactical positioning is impossible to say, as it all could be just a continuation of the usual lies. Nevertheless, a widespread public thaw had begun. Criticisms of Party views, especially those of Révai, started appearing. Satires with political content were written and produced by Kossuth Prize winners who a year earlier toed the line. There was a release of the general fear embodied by the populace, and some public editorials appeared by 1954 on the "new line." Politically, however, Rákosi still had his supporters and the Nagy-aligned "revisionists" would lose key internal Party debates. Underneath it all, the fundamentals were changed little. The near-term result of Stalin's death and Nagy's first ascent to leadership was for the writers to divide into two camps, playing out ideologically the ongoing battles between Rákosi and Nagy. The Party machinery was divided and slowed, but not stopped.

The next critical event was the release in 1954 of hundreds of political prisoners after Nagy's closure of the internment camps. Included in this group were those who had been interred during the worst years of the Ice Age, many of them former militant underground Communists or devoted Stalinists caught up in the Rajk conspiracy, such as János Kádár. Imagine the shocking appearance of dozens of influential professionals, academics, and bureaucrats given up for dead, now prematurely aged or physically decrepit. The charges against them were for mysterious, conspiratorial, anti-Hungarian activities, often with sentences of death or life imprisonment. Many had been arrested by the ÁVH and sentenced through a faked trial. They had been beaten and tortured, starved, or imprisoned in secret ÁVH cells. The survivors knew of many who had been executed by hanging and others who had gone mad, tortured from the physical and mental pain brought on by incredible sadism.[12]

Now they were free. What had happened? The house of cards that was the Rajk conspiracy of five years earlier came falling down. It was suddenly clear that Rajk's trial, all the confessions, the Titoist conspiracy, the public denouncements, Rákosi's sleepless nights, and Révai's attacks on the writers made up one stupendous lie, created for those who had not been imprisoned or murdered. Those who had escaped imprisonment or death felt overwhelming shame and culpability for having actively promoted false beliefs in others' crimes. For writers, this meant their recognition of having fabricated stories to support the torture of innocents. The content of the Déry debate of 1952, some now realized, was that writers had been convinced not to "moralize" about the "truth," just as Révai had told Déry not to moralize in his novel. The truth was now known that the writers had been accomplices to the Rajk conspiracy, that they had falsified for Rákosi's show trial of the century, with Révai, the brilliant essayist, himself the author of Rajk's phony confession. The poems, stories, radio shows, films, and speeches on the Tito-Yugoslav threat were all falsified history, fabricated pieces in the "realization," and the apotheosis of the Lebenslüge. The remorse and personal mortification was overwhelming for the many who had believed, and still believed, in Communist ideals. "My crime," many knew, "was to have believed in yours" (*RM* 268). The split that had existed among the writers under Nagy disappeared with the universal recognition of their complicity in Rákosi's personality cult and the constant falsification of Hungarian life. Party policy for writers had been to celebrate them, and now it was turning against itself, as at this point Communist writers could not easily be imprisoned for political disagreements. Writers finally were not at risk of arrest for criticism, or were willing to take the risk, and through poetry, fiction, and editorial policies, Hungarian society began to open in earnest.

Budapest's changed climate was noticed in Moscow. Nagy was called there in 1955 and asked to self-criticize as a sign that he would toughen up, yet he refused. He was then expelled not just from his position as prime minister but from the government and Party entirely. Sent into retirement though not exile, Nagy was replaced by an associate of Gerő's, András Hegedüs.[13] Nagy's refusal to self-criticize would be the first of many courageous speech acts. He initiated the negation of Stalinist self-criticism, and began the definition of the revolution as one for freedom and truth. At the same time, Moscow brought the Rajk conspiracy and its fabrications back into the orbit of Hungarian politics, and the political oppression became nearly as bad as it was before Nagy assumed power. In the summer of 1955, Khrushchev traveled to Belgrade to improve relations with Tito. Included in the new rapprochement was the need for Rajk's "rehabilitation," as such criticisms of criticisms are called. With the release of the political prisoners, Rajk's ghost was everywhere. In the editorial offices of the largest Budapest newspaper, *Magyar Nemzet* (*Hungarian Nation*), Miklós Gimes, later one of five including Nagy condemned to death in 1958, would ask for "clarification" of Rajk's status, simply an astounding proposal after the previous decade. The government was finding it almost impossible to deny its role.

It took another year, until October 1956, for Rajk's rehabilitation and reburial. In September 1955, the government had completely alienated writers and others by confiscating, through a bureaucratic bungle, the journal of the Writers' Association, *Irodalmi Újság* (*Literature Today*). The journal editor was fired, followed by resignations of Communist members of the governing "presidium" of the Writers' Association. Such was the prelude to an epochal Writers' Association memorandum of November 1955, originating in the association's internal Communist Party, and thus constituting an authentic and self-conscious Party self-criticism. The memorandum protested against censorship, limits on writers' autonomy, the confiscation of *Irodalmi Újság,* the repression of criticism generally, and specific bans like that against Imre Madách's *Tragedy of Man,* the so-called Hungarian *Faust.* It was signed by the "elite of the Hungarian Communist intellectual life, by the favorites of the regime. . . . It was a *Communist* text—and, therefore, it was all the more damaging to Party leadership" (RM 350).[14] The memorandum stressed that even Communist writers could not work under present conditions and demanded rights for non-Communist writers as well. The cynical practice of self-criticism as humiliating contrition was here exploited as valid self-criticism from within, guided by a moral obligation to make possible the right to publish freely and honestly.

In this beginning to the revolution, it was less the text than the deed, another sophisticated speech act intended as a true Party condemnation. Taken out of context, the origin of the text may not seem especially significant. But the memorandum carried tremendous symbolic weight because of its Party provenance and the cynicism of traditional self-criticism. The memorandum was approved by a majority of the Communist membership of the Writers' Association, itself another unheard of event in the Soviet Union or bloc countries. An immediate counteroffensive was launched by Rákosi to obtain retractions, and he largely succeeded. Only eight signatures finally remained of the original fifty-nine, leaving just writers and no longer any film or theater directors or artists. Yet by this time the retractions were moot, with the writers' honest deed beginning to compensate for the orgies of self-criticism of earlier years. News reports circulated in the Western press of the Writers' Association and their protest, though even some Westerners believed the story to be a concoction. Following the memorandum, the winter of 1955 and spring of 1956 marked the heroic age of the Hungarian writers' movement along with the emerging public liberation of writers and journalists from Party truth. The Party transmission belt of lies and propaganda was finally unequivocally destroyed.

The writers and much of the intelligentsia had come full circle. Many now understood that they had been compromised souls like Révai and Kádár. What was next? They were no longer sophisticated historical falsificationists or impersonators of their own false self-criticisms, but rather, real self-critics of the Party to which they belonged. They no longer believed the macabre fantasy of Rajk's Titoist venture, yet they were still part of the Hungarian elite, now leading the country toward a revolt for truth and freedom, as well as for urgent changes in education, economic planning, industrial policy, labor relations, the military, and politics.

Remarkably, the voices for Hungary's manifold needs would also come from within the Party. In February 1956, delivering his secret speech against Stalin to the twentieth Party congress, Khrushchev finally denounced Stalin's "ideological justification of terror" and horrific crimes. There were other means, Khrushchev said, of building a socialist state, and he called for a definitive break with Stalinist methods. Pressure immediately mounted on Rákosi to resign as the realization grew that Hungary's problems did not stem from a single person and his personality cult but were rooted in a thoroughly corrupt system of government and politics.

In March, a group known as the Petőfi Circle was founded under the patronage of Hungary's Communist Party youth organization. The circle's nominal aim was to debate the implications of the twentieth congress for Hungarian intellectual life. Tactically, the Party expectation was that these debates would provide a relief valve for critical sentiments among young intellectuals. But Petőfi Circle leaders, who were also disciples of the ousted Nagy, took Khrushchev's anti-Stalinist sentiments seriously. The Petőfi Circle began a historic series of public debates on economics, philosophy, journalism, history, education, and the arts, addressing the failures of the five-year plans, the falsification of Hungarian history, Soviet distortions to the Hungarian educational system, and so on across the board. First hundreds and then thousands attended the debates, which had to be broadcast to crowds outside the meeting halls. Finally, all the issues, which had been restricted to closed or clandestine meetings and mostly elite groups, were extended to the entire country. The content was assuredly anti-Rákosi and anti-Soviet, but not anti-Communist. Just as the Writers' Association staged their self-criticism from inside the Party, so too were the Petőfi Circle debates immanent critiques of the Party. Whether by accident of history or otherwise, the Hungarian Revolution was a Marxist-Communist one launched through the criticism of Communist elite groups by themselves.

The exquisite logic was lost on no one. The philosophy sessions included the rehabilitation of Lukács, who had been excluded from public affairs since 1949. The widowed Júlia Rajk, in an unprecedented speech now protected by moral outrage, announced still unaddressed questions about László Rajk's show trial and demanded punishment for her husband's murderers. A "press debate" on June 27 drew six thousand people and lasted for nine hours, with Déry speaking on Révai's cultural policy. Although everything about Hungarian Stalinism could justify the most vicious scapegoating, Déry asserted that Hungary's problems were systemic and should not be blamed on individuals. Déry called for a 1956 youth like that of Hungary's earlier war of independence in 1848. At that time, Sándor Petőfi and Pal Vasvari seized the printing press of one of Hungary's leading publishers and declared freedom of the press, just as Hungarians of 1956 were demanding freedom of expression as a prerequisite for reorganizing civil society as well as the country's dilapidated and irrational economic infrastructure. Déry also did not demand dissolution of the Party, and all wanted Nagy returned to government and leadership.

Moscow at last relieved Rákosi of his position as Party first secretary and canceled his membership on the political committee, with Gerő appointed his successor, but only in part because of the obvious public ferment. Rákosi had been planning a final purge of about four hundred individuals secretly listed for mass arrest. He presented his plan to Hungarian Party leaders, who then informed Moscow. A personal visit by Anastas Mikoyan to Budapest led to Rákosi's quick expulsion from the Party, and at last his self-criticism. At this point, who could tell or care what it might mean? As usual, Moscow's interest was not so much with the troublesome Hungarians, as Mikoyan's next visit was to Tito, to whom he could present Rákosi's head as the principal organizer of the anti-Titoist Rajk conspiracy: just another rewrite of history.

With Rákosi's demise and exile to the USSR, the summer of 1956 saw a complete political thaw. The literary journal *Irodalmi Újság* was now openly recognized as a primary forum for critical views. Its circulation rose from 8,000 to 30,000, and other papers and magazines took on the same critical approach. Writers and the nation were finally convinced that there were not two truths in Hungary—a bourgeois and a socialist truth—but authentic categories of truths and lies, and that the elitist theory on which the Party was built only covered up illegal deeds and violence. To speak truthfully and freely meant the possibility

> to think in a non-Marxist way; or to think in a Marxist way even if his thought has not yet figured among the officially proclaimed and obligatory truths. . . . Some of these people, who, for many years, had been spokesmen for falsehoods and lies (whether they knew it or not), felt that now they were in the right at last. . . . For the first time in a decade, the writers were truly popular. After the false popularity that had surrounded them for so long, they enjoyed bathing in the love and unity which pervaded the country. (*RM* 429, 432)[15]

Truth, though, was still barely realized as part of Hungarian life. Júlia Rajk's fight to reinter her husband was at last won with a reburial set by coincidence for October 6, the official day of mourning for thirteen generals executed in the 1848 war of independence. The day was cold, rainy, and windy, yet a mountain of wreaths was piled at Rajk's bier. Júlia Rajk and her young son stood vigil through the day as a hundred thousand mourners paid their respects. Present were many who had conspired against her husband, recalling for some a folk custom referred to as the "ordeal of the bier." When someone was murdered and the murderer was unknown, the village would line up single file and walk by the victim's coffin, whose wounds would open and bleed when the murderer passed by. The ordeal of the bier was just the beginning for Hungary to purge itself of its years of deceit. The Rajk trial, and then Rajk's reburial, brought together all the elements contributing to Hungary's macabre years since the war, as summed up in Ferenc Münnich's speech for Rajk: "He was not allowed to die a great death, a death worthy of heroes, he was killed by sadistic criminals who had crawled into the sun from the stinking swamp of a 'cult of personality.' This swamp was a breeding ground for the falsification of history, for careerism, for contempt for tradition and law" (RM 438).

The Hungarian Revolution began on October 23 and effectively ended thirteen days later, on November 4, with the massive invasion of Soviet tanks and armed personnel. Nagy became prime minister on October 24, and for these brief days, among much else, the country was in control of its own speech. Of October 23, Georg Konrád wrote in his novel *The Loser,*

> You could walk down the street with a flag before this day, too, but only on national holidays, passing by a reviewing stand and cheering the party leaders. It's a natural wonder: within a matter of hours, the populace became a people. . . . Let's tear off the straightjacket of fear, let's speak the truth for a change. Until now we used the lingo of the powerful to lie to one another; now we exchange words as though we were making love, and will not have our sentences approved by the censor enthroned in our heads. We are walking utopias; on a piece of paper we write down what we like and post it on the first tree. A whole system of rhetoric has crumbled; language has rebelled, everyone is a writer, the whole city a bulletin board. The whip can go on cracking, but the horse refuses to pull the cart.[16]

Here this remarkable saga of elites, falsified history, and heroic self-criticism ends for us. The facts are clear. János Kádár, appointed Party head on October 24, disappeared during the night of November 1, to return to Budapest with his Soviet controllers. He would be the titular leader of Hungary until 1988. Nagy and others, including Lukács, took refuge in the Yugoslav embassy, then through a deceit, were abducted and held in Romania. In early 1958, Nagy's secret trial was held in Budapest, with Nagy and his codefendants executed on June 16. Truth teller to the end, Nagy had not self-criticized in 1955 in Moscow, refused to recognize the legitimacy of the Kádár regime after November 4, 1956, and heroically told his judges in his final words, "I do not appeal to the court for clemency."[17] Many others were tried and given harsh sentences, including Déry. An important voice for writers during and after the revolution, Déry was sentenced to nine years imprisonment in 1957, then was amnestied in 1960. Lukács, who had been minister of education under Nagy during the uprising, returned to Budapest and was not imprisoned. Political trials continued until as late as spring 1959, with tens of thousands subject to investigation, and several hundred executed for their roles in the revolution. The numbers are comparable to the worst years of the Rákosi regime.[18] The revolution was reinterpreted as a counterrevolution, freedom fighters became criminals perpetrating sabotage, and so on, thus reconstructing Hungarian history one more time. Writers and intellectuals were strictly monitored, and the country fell into catatonia, with few even speaking of the revolution or its defeat. This grim scenario of self-censorship and Orwellian control was what the surviving revolutionaries unknowingly had before them when they radioed the world for help during the dawn hours of November 4, 1956, as Soviet tanks pummeled town and countryside:

> This is the Association of Hungarian Writers speaking to all writers, scientists, all writers' associations, academies, and scientific unions of the world. We turn to leaders of intellectual life in all countries. Our time is limited. You all know the facts. There is no need to expand on them. *Help Hungary!* Help the Hungarian writers, scientists, workers, peasants and intelligentsia. *Help! Help! Help!*[19]

Lakatos's Road to 1956

Where was Imre Lakatos during the Hungarian nightmare? What activities was he involved in and how does his experience compare to that of other Hungarians? His intellectual profile is much like that of similarly talented Hungarians: militant dedication and zealous support of Stalinism followed by an equally strong volte-face against the tyrannical truth of falsity. Far from being an ordinary Hungarian, the Stalinist Lakatos was involved with luminaries of the Hungarian intellectual world as well as powerful members of Rákosi's regime and some of their intrigues. This is truly Lakatos's "external history": the material personality, conditions, and events associated with the ideas explored so far. Lakatos's story is as interesting as they get, and worth knowing by any of his readers. Just how this information should be integrated with the "internal," Popperian, "third world" philosophy and history of ideas that is the subject of this book is taken up in the next and final section.

Lakatos' was born Imre Lipsitz in Debrecen, Hungary on November 9, 1922.[20] His father was a wine merchant who left his family early on, forcing Lakatos and his mother to move in with an uncle. As a high school student, and later, Lakatos excelled at mathematics, and during 1937–1938, "Imre Lipsitz" appears several times in the prestigious national high school mathematics journal, *Középiskolai Matematikai és Fizikai Lapok* (*Mathematical and Physical Journal for High Schools*), where the likes of John von Neumann and Paul Erdős would first make their mark in the mathematical world.[21] He studied mathematics, physics, and philosophy at Debrecen University, where one of his teachers, and later a political confrere, was the philologist and historian of Greek mathematics Árpád Szabó, whose views on the role of dialectic in the origins of mathematical proof in ancient Greece complement Lakatos's account for modern mathematics.[22] Lipsitz became Lakatos during the war, and for a time Molnár before he was Lakatos. He graduated from Debrecen in 1944, having joined with others to form an underground anti-Nazi Communist cell. Included in this group was Éva Révész, whom Lakatos would marry in 1947. Lakatos's mother, uncle, and grandmother would die in Auschwitz, while his father survived the war and would later correspond from Australia with Lakatos in England.

In 1947, Lakatos completed a dissertation at Debrecen, titled "*A természettudományos fogalomalkotás szociológiájáról*" ("On the Sociology of Concept Building in the Natural Sciences"). No copy is known to exist; indeed, Lakatos himself may have stolen the library copy before leaving Hungary. László Ropolyi has attempted to reconstruct the content of the dissertation based on a series of contemporary articles by Lakatos and the comments of his dissertation adviser, Sándor Karácsony of Debrecen, indicating that the content of the articles and dissertation is similar. Remarkably, the approach appears to be exactly opposite to the ideas of semiautonomous scientific criticism developed in the methodology of scientific research programmes. Ropolyi conjectures that the dissertation "probably presented a fundamentally Lukácsian analysis of the history of natural sciences, first of all physics, based on Lukács' *History and Class Consciousness*."[23] As described by Gábor Kutrovátz in his translation of Lakatos's 1947 article "*Modern fizik, modern társadalom*" ("Modern Physics, Modern Society"),

> Instead of focusing on the immanent development of science, he decides to look for explanations outside of science. He emphasizes the indispensable role of sociological and economic influences on scientific concept building, and he concludes that the worldview of a given scientific age or community is nothing more than a historical category. . . . [In "Modern Physics, Modern Society"] we are given a deeper (Marxist and Lukácsian contra Hegelian) analysis of the "dialectical structure" of the modern scientific view determined by social relations and motions. And if we imagine that we go further in this direction, then we must be very close to the text of the lost dissertation.[24]

The influence of Lukács is evident, for example, in Lakatos's appeal to Lukács's ideas on "commodity fetishism" from *History and Class Consciousness.*

After the war, Lakatos would move quickly into political work, and he is remembered by contemporaries as a ruthless and charismatic leader. "He calculated everything, in his eyes everyone was a means," recalls sociologist István Márkus; the widow of psychologist Ferenc Mérei, one of Lakatos's best friends, says Lakatos was "diabolically clever, a genius"; historian Béla Köpeczi describes him as "a fanatical Communist who believed the end justified the means."[25] Lakatos was recognized as a superb debater, but at times, he used the ruse of citing an authority known to his opponent when he was simply fabricating the supposed fact impromptu. A story about Lakatos's wartime underground cell that circulated for years in various forms, and has finally been documented through archival research, involved a nineteen-year-old woman named Éva Izsák. Izsák participated in underground work and came to the Nagyvárad area, where Lakatos and his group hid out in various safe houses. Here, Izsák joined up with Lakatos and Révész, and while she was politically correct, she had difficulty keeping covert. Perceiving Izsák as a threat to the others' safety, Lakatos proposed the solution that she should commit suicide. At a group meeting, the terrified Izsák pleaded that another solution be found. Lakatos called for a vote, which went unanimously against Izsák; even her lover, one Alfonz Weisz, supported the suicide option. The vote taken, Lakatos is said to have delivered a lecture to Izsák and the group that correct theory, once deduced, always had to be translated into practice. Shortly thereafter, Izsák was escorted to a wooded area around Debrecen and provided with a cyanide drink. Izsák consumed the drink and quickly died; her overcoat was appropriated by Lakatos's future wife, Éva Révész. The events, perhaps including a hanging rather than poisoning, or with other details changed, were alleged by Lukács's student Ferenc Fehér to be recounted by Lakatos with relish.[26]

Lakatos became a cog in the Soviet-Hungarian power machine after the war. He worked closely with Lukács and read his major works, including *History and Class Consciousness,* and later in the 1950s, *The Destruction of Reason.*[27] Lakatos worked for the Ministry of Education between 1945 and 1948, and was an active writer supporting the hard Party line against the liberal factions to be demolished by Rákosi. During the 1947–1948 academic year, Lakatos dedicated himself to helping destroy the distinguished Eötvös College, targeted by the Communists because of its resistance to transforming itself into an indoctrination tool like the recently established Györffy College. Eötvös College was founded in 1895 as a teachers' school. It supported the education of a pedagogical elite, including many of Hungary's leading educators and cultural figures, with preference given to students requiring significant financial support. Györffy College was intended for students coming from the peasantry and early on was dominated by militant student Communists, even during the war.

The detail is telling, not just because of Lakatos's later dedication to philosophical pedagogy but to what became of Eötvös College due to Lakatos's intervention. Through pressure brought by Révai, Lakatos, in spite of having already graduated from Debrecen, was allowed to enroll at Eötvös College. He had expertise on educational issues from his Party and ministry work, and therefore was a perfect choice for disrupting college activities from within. Lakatos's enrollment was forced by the Party on the director, Dezső Keresztury, and once at the college, Lakatos was able to help Communist students undermine the school, paralleling Rákosi's destruction of the Smallholder's Party outside. Keresztury would write in 1989 that "Imre Lakatos' entry marked the beginning of the College's end. More precisely, he was the emblem of ruin, the instrument of great powers."[28] Keresztury was also the minister of education at the time, and Lakatos's campaign against him was systematic and intense. In articles and speeches, Lakatos polemicized against the Eötvös curriculum, intimating that the college provided no significant resistance against the fascists and was unsympathetic to the aims of a people's democracy. The college's goals, argued the young Lakatos, were fundamentally elitist, and thus, unsupportive of the great social transformation taking place. The ultimate problem, though, was college leadership. Lakatos suggested in an article, "Eötvös College-Györffy College: Eötvös College in the Balance," that if Keresztury could not institute radical reforms, which appeared unlikely, he might as well be the college's *last* director. Intimidated by the attack, and out of fear and his dislike for Lakatos, Keresztury did not reply to the article. Some Eötvös students, themselves Communists, answered Lakatos in print, but they were no match. Nor could they withstand the treacherous political machine of which he was a part. Keresztury resigned his post, and in 1950, Révai disbanded Eötvös College. Lakatos "subverted the College," Keresztury said in 1992, "and managed to bring about its demise." Though Lakatos's influence should not be overestimated, such was the talented and charismatic young man described in an Eötvös College student yearbook as "Mephistopheles" and "a slinking wolf."[29]

As a reward, Lakatos was sent to study in Moscow. But he was arrested there by Soviet police when a Romanian Communist denounced him for his role in Izsák's suicide. "Because of his murky past and incorrect attitude," as an internal Party document put it, Lakatos was recalled to Hungary.[30] He brought with him pages of material culled from Révai's articles and speeches, who is said to have hated Lakatos. Back in Budapest in April 1950, Lakatos compiled his Révai notes into a memorandum titled "Révai's Crimes," delivered it to the Party Center, and was arrested the same evening. Lakatos was expelled from the Party and taken into custody by the ÁVH, possibly because of a plan to denounce Révai as an imperialist agent based on Lakatos's research: Révai, for Lakatos, was not Stalinist enough.[31] Lakatos ended up at the Recsk forced labor camp, where he remained for over three years. Recsk was a heinous institution for its approximately one thousand inmates. Day work consisted of hard labor such as road building and stone quarrying, with those not meeting quotas punished by reductions in what was already a starvation diet. Prisoners were brutalized and many inevitably were in poor health. Lakatos nonetheless may still have been cooperating at Recsk with the ÁVH.[32] In any case, Lakatos survived Recsk until it was closed by the Nagy government in 1953. Lakatos maintained the Party line after his release, arguing that Rákosi must have had good reasons for his actions. He continued, as he had for years, to spy and inform on others, including his mentor Árpád Szabó. Recall that after the prisoners' release, the writers and others were opposed in two camps supporting Nagy and Rákosi. Lakatos was still a staunch Rákosi backer.

Lakatos obtained a position in 1954 as a librarian and translator at the Hungarian Academy of Sciences' Institute of Mathematical Research through Alfréd Rényi, a world-class mathematician and pedagogue. Like Szabó, Rényi saw mathematical technique as continuous with Platonic-Socratic dialectic.[33] Here Lakatos had the opportunity, as did other fortunate individuals throughout the Soviet bloc, to read prohibited authors, including Popper and Orwell. By the fall of 1956, Lakatos appears finally to have changed his views, just as many within the writers' movement took that long to rethink theirs, even after hearing the horror stories of released prisoners. Lakatos participated in the education section of the Petőfi Circle debates in October. Along with many others—gatherings and meetings continued all day—Lakatos delivered a speech against the regime commensurate with his usual skills, but now targeted against the inane ideology of sanctioned Soviet science, including Lysenko genetics.[34] On October 30, Lakatos drafted a position statement for the newly organized National Committee of the Hungarian Academy of Sciences:

> The National Committee of the Hungarian Academy of Sciences takes its stand with true freedom of science. Only his own scientific conscience may guide the scholar. We demand that every scientific convic-

> tion be allowed to be freely expressed in written and oral form, in universities, in scientific institutes, and in other public forums; free from every power restraint and moral pressure. Without delay a general meeting of the Hungarian Academy of Sciences must be convened. This meeting will be competent to rehabilitate unjustly neglected and oppressed scholars and scientific trends, and to liberate Hungarian science from the shackles of Stalinism.[35]

Lakatos fled Budapest on November 25 with his in-laws and second wife, from whom he was later divorced in England. His brother-in-law, a student and Petőfi Circle activist, had committed suicide on November 2 after being confronted by two former friends who had denounced him to the police. Lakatos and his family crossed the border and went to Vienna, and then continued on to Cambridge, where Lakatos joined many younger and less worldly students in the study of philosophy and the history of science and mathematics. Popper would become his mentor at the London School of Economics, where Lakatos taught until his death in 1974. The story of Izsák's suicide and Lakatos's responsibility was almost forgotten. In 1989, Izsák's sister, now Mária Zimán, privately published her account in Israel of the tragedy and Lakatos's role.[36] Imre Lakatos, born Imre Lipsitz, the great critic of historical falsification, the brilliant composer of historical rational reconstructions in science and mathematics, was quite competent in suppressing his own footnote to the Holocaust.

In the context of the writers' movement and Hungarian Stalinism, Lakatos was a representative Hungarian intellectual of the postwar years to 1956. Talented and industrious, like many others he used his intellectual gifts to support the Stalinist state. Though one should be cautious about moralizing about years of almost unimaginable danger and paranoia, Lakatos was clearly treacherous and able to carry out diabolical plans as needed for his superiors. A henchman of Révai's and then his enemy, he was a Party operative at the highest ranks and a classic victim of the purges. He suffered in the Recsk concentration camp, but would still support Rákosi. By the time of the Petőfi Circle debates, Lakatos emerged as a superb writer and speaker, his considerable skills finally turned against the regime. This was the émigré who created the dual philosophy of science and mathematics described in this book.

What should be made of this person and his ideas? The rules of Lakatos's dissembling philosophy do not dismiss a biographical explanation, but they resist it. Wherever he could in the history of science or mathematics, Lakatos wanted to explain its changes "internally," meaning with reference to largely autonomous relations among ideas with no special dependence on "external" social, psychological, or political history. My opinion is that it is most consistent to respect that directive and make little of Lakatos's personal past in interpreting his philosophical work. Hungary's road to 1956, including Lakatos's as a characteristic example, is supremely relevant to Lakatos's ideas, but *he* in particular can be left out. Hungary's history, rather, is one worth interpreting through Lakatos's work, and provides some closure on this strange man's legacy.

No Philosophy Is Innocent

Lakatos's philosophy can be interpreted against three different possible histories in order to draw some broad philosophical morals going beyond the detailed content of his work. First is the received history of Lakatos's philosophical work in England as that of the émigré author who "gave up" his Hegelian past to become Popper's ardent disciple and then critic. In this "English" history of ideas, assumed by most of Lakatos's readers, Lakatos's work is a new start from Popper and a well-known voice in the Anglo-American philosophy of science debates of the 1960s and 1970s. Hegelian-Marxism, Hegel, Goethe, Lukács, and the rest have a superficial role in this history.

Second is the history of Hegelian-Marxist ideas that suffuse Lakatos's work. Here one might imagine, as a heuristic, a Lakatos who never left Hungary and wanted to develop a historicist Hegelian-Marxist account of science and mathematics, using the techniques and concepts that we have seen occur prominently in Hegel, Marx, and Lukács. That philosophy is one that provides a corrective to Hungarian Marxist-Stalinism in particular through Lakatos's coordination of science, historical writing, criticism, and the role of elites. Indeed, Lakatos's 1956 criticisms of Soviet-dictated science offer a perfect starting point for applying his ideas. Even the scientific status of Marx's economics itself, as we saw, can be carefully assessed in that possible world. That contribution is objectively present in Lakatos's writing, whether he intended it or not, and regardless of his repudiations of his Hegelian past. Table 1 (p. 294) summarizes the main concepts, described earlier in this book, transversing these first two histories.

The third history of ideas is that just surveyed: the 1956 Hungarian Revolution, from Rajk to Révai to Déry to Lukács and others. The relation between the English and Hegelian Lakatos represents Lakatos's philosophical work; the refracting history of 1956 completes the circle.

The primary correspondences between Lakatos's philosophy, meaning the bildungsphilosophie of *Proofs* and the methodology of scientific research programmes with its reflexive historiography, *and* the history of Hungary's 1956 uprising include the following.

1. Falsification and the reconstruction of history. Historiography is the center of Lakatos's philosophy with a primary thesis being the inevitability of normative reconstructions of the scientific past. These reconstructions will be theory-laden by some normative methodological theory whose "internal" history can radically diverge from "actual history." Actual history, too, is not grounded in apodictic historical facts, and multiple histories can be critically compared against one other. In Hungary, an extreme abuse of historical narrative was apparent. The Rajk conspiracy was a complete fabrication, a "realization" as it was called, rather than a "reconstruction," with no "comparison" against actual history. Déry could not compare actual history against Révai's humiliating "rewriting" of his novel *The Answer,* nor did Hungarian writers manage for years to compare Hungarian reality against the hundreds of poems, stories, and articles produced by them in support of Rákosi and the Party. The pointed or slanted rewrites of coerced confessions were—like the stereotyped gestalts of *Proofs,* or Hegel's *Phenomenology,* or Marx's *Capital*—the ÁVH's distillation of the allegedly essential aspects of conspiratorial history. Hungarian writers, social scientists, film directors, and others rewrote Hungary's twentieth-century history on a day-to-day basis using the "normative" theories encouraged, for example by Révai, to correctly represent the trajectory of the working class and the triumph of the Communist Party, or Révai's rational reconstruction of the way history should have gone. Falsified history included mundane economic facts, political slander, revisionist historical interpretation, fictionalized history, and conspiratorial fabrication. Ordinary falsification of, say, production achievements was never compared against the real amounts or quality of produced goods. Lasting about a decade, Hungarian literature, journalism, history, and education was historical reconstruction gone mad: a total falsification of the past in the service of an insane political regime.

It was this collection of falsifying practices, in the name of Party truth, that the writers' movement recognized after Stalin's death as their collective responsibility, and against which they defined their opposing stand for truth. The Hungarian Revolution was, in significant part, a revolution against these perverse, power-distorting forms of historical reconstruction. Lakatos's central technique of rational reconstruction, minus any critical process for the assessment of historical claims—and accompanied by an exact inversion of Lakatos's philosophy in that *who* represents a given view is primary rather

Table 1. Hegelian Themes in Lakatos's Work.

Concept or Technique Appearing in Lakatos	*Relevant Hegelian or Marxist Equivalent or Analogue*
Falsification and learning through error (introduction and chapter 1)	Bildung and negative progress through criticism
Biogenetic method of *P&R* and the caricatured shapes of mathematical method (chapter 1)	The historiographical method of *Phs*, which Lukács placed in descent from the bildungsroman and Goethe's philosophy of learning
Reordering of history of mathematics according to the role played in contemporary mathematics (chapters 1 and 5)	Used in *Phs*, and by Marx through his "order of presentation" versus "historical order"
Commentary via footnote apparatus (introduction)	Used similarly in *Capital*, but not in Hegel
Heuristic as means for making method exoteric (chapters 1, 2, and 5)	Same motivation in *Phs* for philosophy as presented in its preface
Antifoundational view of knowledge coordinated with historicism (introduction and chapters 1, 3, and 7)	Hegel's transformation of Pyrrhonian skepticism in his historiographical method in *Phs*
Immanent critique of Popper and historicization of methodological categories (chapters 7, 9, and 10)	Marx's critique of political economy and historicization of economic categories
Hindsight in research programmes (chapter 8)	Hegel's retrospective rationality and "Minerva's owl"
Contradictions in research programmes (chapter 8)	Contradictions in the natural or social worlds rather than statements or theories
Theory-laden history; history of discourse versus history of events (history$_1$ and history$_2$); internal and external history (chapter 9)	Same in Hegel's historiography
Rational reconstruction of history using normative philosophical categories (chapter 9)	Hegel's various philosophical histories of art, political history, and religion
Falsification of historical reconstructions by actual history (chapter 9)	Marx's "measuring Hegel by his own yardstick"
False awareness of scientists of their own methodological practices (chapter 9)	False consciousness of historical actors in Hegel, and classes in Marx
Elites as sources of value judgments on scientific method (chapter 9)	Hegel's world historical individuals, and Marx and Lukács's vanguard classes
Methodological honesty (chapters 7, 9, and 11)	Methodological honesty for Marx in political economy

Table 1. *Continued*

Concept or Technique Appearing in Lakatos	*Relevant Hegelian or Marxist Equivalent or Analogue*
Antipathy toward nondiscursive philosophies, "bad" elites; identification of irrationalism in antihistorical method (chapters 1, 5, and 12)	Hegel's polemic in preface to *PhS*, and Lukács's account of epistemological elitism in *DR*
Popper's objective knowledge and language as semiautonomous (chapter 5)	Hegel's Geist as expressed in language and speech acts, but absent in Marx's social theory
Self-application of the methodology of scientific research programmes to create a changing logic (chapters 9 and 10)	Hegel's attempts to describe a theory that is its own metatheory; Marx's simultaneous presentation and criticism of the system of political economy; Lukács's application of Marxism to itself to demonstrate the changing function of historical materialism

than *what* that view is—appears in Hungarian history as a mechanism of totalitarian control and the cause of self-conscious revolt.

2. *Elitism and the personality cult.* Two types of elites play important roles in both Lakatos's philosophy and the Hungarian Revolution. First is the positive role of writers as bona fide elite leaders of Hungarian life and politics. Writers were not forced on the Hungarian public as were other components of the Soviet model. A long-standing literary and poetic tradition accepted the expression of progressive and radical political views, and the active participation of writers in political activity. As such, writers were acknowledged as representatives of popular views and movements. Hungarian literature was not an autonomous cultural sphere "reflecting" political life but was directly political itself. Accompanying that privileged position was a moral responsibility to protect writers' influence from abuse. As we saw, many Hungarian writers, with the exception of heroes like Déry, defiled their role as a privileged elite through their writing and the falsification of Hungarian life for an immoral political regime. The elite status of writers and others, as effective sources of perceived truth, was leveraged in turn by the Party's monopolization of media control. For Lakatos, elites play an analogous part in providing *initial* accepted assessments of good and bad, or progressive and degenerating, science. Their judgments are provisionally accepted as speaking for scientific consciousness, as it were. Lakatos's elites are essential in the process of reconstructing science's past, and the scientific

elite has ethical responsibilities analogous to those of other intellectuals as sources of specialized value judgments. Again, for Lakatos, these views are discursively criticizable through means that did not exist in Stalinist Hungary.

The second type of elite found in Lakatos's philosophy and Stalinist Hungary is a perverse form of the bona fide elites. These elites embody the infallible sources of knowledge represented by Rákosi and the Communist Party hierarchy, the latter being a direct descendant of the vanguards of historical consciousness in Hegel, Marx, Lenin, and the young Lukács. Rákosi and the Party were the antithesis of discursive objectivity promoted by Lakatos in *Proofs* and Hegel in the *Phenomenology.* Indeed, Rákosi's role as center of the personality cult and privileged source of political truth is exactly the romantic subjectivity that Hegel saw in Schelling, and against which the *Phenomenology* was in part written. The personality cult also shares its epistemology with the implicit authoritarianism Lakatos portrays in the Euclidean style of mathematical pedagogy, as hiding a proof's historical genesis and heuristic structure.[37] The personality cult, used as a means of psychological and social intimidation throughout Hungarian society, realized a version of the romantic, subjective, and intuitive truth rejected by Hegel, and the antihistorical, dogmatic, and intuitive sources of elitist truth rejected by Lakatos. The personality cult, as a sinister form of elitism, co-opted writers as cultural elites to serve Party ends through their zealous falsification of Hungarian life and history.

3. Self-criticism and "dialectic." A central feature of the 1956 revolution was that major oppositional events were self-consciously moments of Communist self-criticism, including the 1955 Writers' memorandum and the 1956 Petőfi Circle debates. Part of Hungarian heroism was that the criticisms of the political system came from inside, as an immanent critique of the regime by its creators. The revolution was not exactly anti-Communist.[38] It was an anti-Stalinist insurrection led by Communists that spread to the whole nation. The exquisite beauty of the revolution was also that self-criticism as a speech act before 1956 had been the modus operandi of fake public contrition and submission to power. The Hegelian-Marxist idea of criticism building knowledge through cumulative solutions to internal contradictions was reduced under Stalinism to a formulaic ritual of confession and tactical subtleties. That anything like reason is here in control is a cruel joke of history. The Hegelian idea of Bildung and self-formative rationality, expressed schematically through the simple skeptical trope of reflexive criticism, falls from its philosophical and literary heights onto one of history's dunghills. Part of the self-critical learning process is the heuristic of identifying the exemplary ideas of the object of criticism and their deficiencies; or as Hegel advised, utilizing a criterion of truth from within a shape of knowledge.[39] In Hungary, that meant a ritualistic acknowledgment of one's errors, meaning what one was ordered to retract; such was the Stalinist fate of Hegel's great philosophical pedagogy as well as a theory of error and learning shared by Goethe. In Lakatos, as in Lukács, the idea of self-criticism was a heuristic to maintain a consistent historical perspective on, respectively, either the philosophy of science or Marxism. In both cases, the skeptical trope guards against mistaking temporal truths for eternal ones, and helps uncover contemporary anomalies or contradictions. This quasi-logical method is a heuristic for understanding change. Under Stalinism, the "change" was for you and your contradictions to shut up and maintain Party discipline, or else. Just as the Hungarian narration of history had degenerated into mendacity, dialectic had degenerated into a perverse political ceremony devoid of serious intellectual content.

Such are the principal affinities between Lakatos's philosophical work and the Hungarian Lebenslüge that transformed itself into a tragic revolt. The Hungarian experience shows why the "second" Hegelian-Marxist history of ideas just discussed is useful: that account explains the pathology of falsification and personality cult, and characterizes the status of physical and economic science, while retaining desirable features of Hegelian-Marxist historicism and criticism. Lakatos's implicit Hegelian-Marxism is a *perfect* response to the degeneration of Marxism and its Hegelian roots under Stalinism. At the same time, a skeptical philosophical problem can also be raised about each of our "correspondences"—historical falsification, elites, and immanent criticism—namely, whether the perverse Stalinist version is "ruled out" by its Lakatosian-Hegelian rationalist twin. That is: Can the perverse falsifications of Hungarian Stalinism and Lakatos's rational reconstructions be distinguished *in principle?* Can Lakatos's needed identification of some elites be prevented from degenerating into authoritarianism? Can the pathetic ritual of Stalinist self-criticism be excluded from its numerous historical shapes, whether in ancient skepticism and the peritrope, or Hegel's intricate phenomenological method, or the variants found in Marx, Lukács, and Lakatos? What prevents reason from becoming demagoguery? I do not want to denigrate Lakatos's achievement of implicitly isolating and addressing the roots of Stalinist epistemology. But neither should it speedily be assumed that Lakatos has driven a stake into Stalin's heart.

Orwell's classic novel *1984* provides some additional perspective on how the three major features of Stalinist culture extracted here fit together.[40] The essential ingredients of Stalinist language and culture that Orwell described in 1949 are all found in Hungary. Orwell's Big Brother is, of course, Stalin himself, or whoever happens to be the center of the personality cult, like Rákosi. Newspeak is the stultifying, simplifying language of the Comintern that invaded and took over Hungarian, but whose vocabulary, Orwell writes, also included many words, like Hegel's aufheben, with "two mutually contradictory meanings." Orwell's infamous Doublethink just is dialectic, or as Orwell comments, "the power of holding two contradictory beliefs in one's mind simultaneously, and accepting both of them."[41] It is a process of "mental cheating" that facilitates the "continuous alteration of the past, made possible by the system of thought which really embraces all the rest."[42] Doublethink, though a supposed historical method, paradoxically allows Big Brother "to arrest the course of history" because history is constantly recreated to serve totalitarian ends. This reconstruction, rewriting, and falsification of history is at the heart of *1984.* That is even the occupation of the book's protagonist and main victim, Winston Smith, who knew that falsification of history was essential to maintaining Big Brother's power through infallibility:

> There is the need for an unwearying, moment-to-moment flexibility in the treatment of facts. . . . By far the most important reason for the readjustment of the past is the need to safeguard the infallibility of the Party. . . . For to change one's mind, or even one's policy, is a confession of weakness. . . . Thus history is continuously rewritten. This day-to-day falsification of the past, carried out by the Ministry of Truth, is as necessary to the stability of the regime as the work of repression and espionage carried out by the Ministry of Love. . . . Every record has been destroyed or falsified, every book has been rewritten, every picture has been repainted, every statue and street and building has been renamed, every date has been altered. And that process is continuing day by day and minute by minute. History has stopped.[43]

Orwell captures the dependencies among propaganda, ordinary language, and historical writing under Stalinism and its personality cult as a perverted, doublethinking descendant of Marxism. Language and communication effectively disappear in Marxian social theory, as pointed out by Habermas, but now they reappear as the medium of totalitarian control whose handmaiden is its topsy-turvy double: an exceptional shape of spirit, a gestalt of

Marxist-totalitarian consciousness and form of life, not because of its mode of economic production but that of communication and control. Its Hegelian Geist appears through propaganda techniques originating in German philosophy and social theory. Language and Doublespeak were means of deliberate control, and the expression of power and submission. The foundation supplied by language for rationality in Hegel, Lakatos, and Popper supports instead the irrationality it was intended to exclude. Orwell isolated historical rewriting, personality cult, and pseudodialectic as features of Stalinist totalitarianism, all of which appear in Lakatos's philosophy in forms to combat their irrationalist twins. The historical fact that Hungary was a Stalinist culture par excellence now is embellished by Lakatos's critical analysis of the epistemology of a Stalinist closed society. The role Orwell gives to language appears in Lakatos's historiographical perspective as well. His histories are in "history$_2$," meaning, as he said, "a theory- and value-laden reconstruction of history$_1$" (*HS* 121), the set of historical events, while under Stalinism history$_2$ *becomes* history$_1$ as a means of social control.

Unlike *1984*, Hungary had many heroes and a moment of freedom created by the subversion of Stalinism. Postwar Hungary was a nightmare of dissembled speech and writing for everyone, but missing from Orwell's Stalinism are the survival skills practiced by many within this world. Kádár, for example, after his release from prison, "behaved as if he sympathized and fully agreed with the movement of the writers; he encouraged them by secret signals and messages," and in the fall of 1955, "writers and journalists pursued their assaults through allusions well understood by the readers" (*RM* 436).[44] The Lebenslüge, in this way, also was turned on its head through dissemblance and cunning. Rationality temporarily triumphed because writers and poets used the same means to invert the Stalinist nightmare that they had used to construct it.

In this sorcery, through which Stalinist wizards became life-giving healers, the actions of the good poets of 1956 are not always easy to distinguish from their dark selves of earlier years. Their words can be poison or cure, like the ambiguous *pharmakon* of ancient Greece. When Lukács writes in the 1957 preface to *The Meaning of Contemporary Realism* that he can finally discuss certain issues "no longer in 'Aesopean language,'" he is not granting himself any special distinction.[45] The dialectics of covert communication became part of the social semantics for a world in which ambiguous language was so charged with power, or submission to power, that extraordinary communicative techniques were necessarily developed to combat the regime. "All Hungarian Communists are liars" spoken by Communists was not paradoxical but true. And like Lukács's correct account of Marx's use of Hegel, Lakatos's relationship to Popper is similarly not just an example of Hegelian criticism but criticism that may be taken as implicit, disguised, or covert.

Now let's deploy the third history and elements making up the epistemology of the Hungarian Revolution—falsification and the quest for truth, personality cult and elites, dissembling and betrayal, and self-criticism—against the first two histories. This Hungarian history is a case of a world the "English" Lakatos wants to portray as a real danger of irrationalism: a closed society with no role for the critical comparison of histories because of its domination by authoritarian elites. No Hegelian or other special reading of Lakatos is needed to see that. But now let's attend to the second history of implicit Hegelian-Marxism, as summarized in table 1 above. This history provides a criticism of Stalinist Hungary, with explanations of scientific and mathematical method, several of them associated even with specifics of Lukács's Western Marxism. For example, Lakatos explores how Marx's "order of presentation" versus the "order of history" functions in the history of mathematics; Lakatos provides a meaning for history as "science," and Marxian economics as science, albeit a degenerating scientific research programme; he gives a role for elites and a process for criticizing their judgments to protect against authoritarian hazards; and he offers a self-critical historiographical theory. Lakatos's work, therefore, is a complete historical philosophy of science, mathematics, and criticism that answers, in the Hegelian-Marxist tradition, philosophical problems raised by Hungarian Stalinism and the 1956 Revolution. The third history, of Hungary, then is an object of opprobrium vis-à-vis this English history *and* the Hegelian-Marxist history, too.

But the third Hungarian history is more than a world whose pathologies just happen to be described by either the English Lakatos or his Hegelian-Marxist twin. Negative features of Hungary 1956 appear in Lakatos as normatively desirable methods. A main practice of Hungarian Communist writers was the falsification of history, yet even for the English Lakatos it is clear that all histories are falsified and guided implicitly or explicitly by some normative theory. The falsifications of Hungarian writers, then, are a perverse extension of a practice that is intrinsically part of a historicist conception of scientific criticism. Stalinist self-criticism also cannot be completely rejected, even for the English Lakatos, since he promotes the trope of self-application to develop his historiographical theory and a changing logic of scientific discovery. Elites as well are not completely rejected; rather, their roles are tempered and open to criticism. Central vices of Stalinist Hungary, as suggested here all along, appear as virtues in the English and Hegelian-Marxist Lakatos. Even the covert subterfuges of Hungarian Communists show that the covertness appearing in Lakatos's work expresses a key feature of the rational triumph of Hungarian writers and intellectuals. Lakatos's covert Hegelianism within Popper's world demonstrates the virtuous role of intellectual dissemblance and cunning in the Hungarian Revolution. The malevolent third history of Stalinist Hungary contains much of what is commendable in Lakatos's philosophy of science and mathematics.

As such, the relation of correspondence between these three histories is like that of the "caricature" that Lakatos himself describes between alternate histories of science: Lakatos's "favorite—and by now well-worn—joke that history of science is frequently a caricature of its rational reconstructions; that rational reconstructions are frequently caricatures of actual history; and that some histories of science are caricatures both of actual history, and of its rational reconstructions" (*HS* 138). Each of the three histories above is a caricature of the other in that each represents or characterizes conceptions of history, criticism, and method that are positively endorsed or present in the others, and conceptions that are repudiated or even disdained from the alternate perspective as well. This is the inverted world of Lakatos's philosophy and the Hungarian Revolution of 1956.

Of course, Hungarian Stalinism as pathological may "obviously" be rejected. But how is that done with philosophical rigor without rejecting the heroic transformation *out* of Stalinism via self-criticism, not to mention cunning and dissemblance, made by Hungarian intellectuals? Surely, we reject Rákosi's personality cult, but where is the line drawn in acknowledging the leading role of Hungarian writers as elites and their part in *promoting* the revolt? Neither Lakatos nor anyone else has some more fundamental moral and epistemological standard by which to make such distinctions. The value-laden interpretation of caricature as mendacious representation is intrinsic to Lakatos's historicism as a potentially nefarious transformation of ideas.

Take, for example, Lakatos's ingenious use of his footnote apparatus to self-criticize his rational reconstructions by comparing so-called actual history with his rational reconstructions: rightfully, this idea provides a critical extension of Hegelian historiography in the spirit of Marx's "measuring Hegel by his own yardstick," as put by Lukács, and was drastically missing from Hungarian historiography. Orwell in *1984*, however, sees the dark side of Lakatos's intricate practice of historical rewrites-plus-criticism as one grounded in doublethink: "To tell deliberate lies while genuinely believing in them, to forget any fact that has become incorrect, and then, when it becomes necessary again, to draw it back from oblivion for just so long as it is needed . . . all this is indispensably necessary"—necessary, that is, for halting history and maintaining authoritarian control.[46] Such is a vice built into the many virtues of Lakatos's historical reason and his clever historiography. Orwell's Winston Smith is pessimistic about determining truth in history, as Winston knew that "all history was a palimpsest, scraped clean and re-inscribed exactly as often as was necessary. In no case would it have been possible, once the deed was done, to prove that any falsification had taken place."[47] In philosophical jargon, historical reconstructions are fallible and theory-laden. Orwell is not so much of a skeptic, yet falsification of history in Lakatos's worlds cannot be "proved" any more than can scientific conjectures. Hence, truth-telling practices have no inherent guarantees against falsificationist pathologies either, and both critical and doublethink sides can appear through Lakatos's rational reconstructions.

A last Hegelian idea can now be introduced to consolidate this notion of virtues become vices and then virtues again. Hegel's description of the relationship between vice and virtue, and representation and caricature, is summed up in his notion of the *verkehrte Welt,* or "inverted world," making non-Orwellian use of the dual meanings of verkehrte as inverted, turned around, or topsy-turvy, *and* perverse. Moral judgments and concepts, such as justice and punishment, are ones that for Hegel need to be understood as containing the internal oppositions of an ideal—say, abstract justice—and its perverse opposite—say, punishment by execution that "repeats" the criminal act as the completion of justice. The punishment can be justified as reasonable and appropriate to the misdeed, while it may also become a grotesque parody of justice itself.[48] For Hegel, the concepts of crime and punishment do not mirror one another as "opposites" but inherently depend on one another in the social acts realizing them, and jointly give meaning to guilt, shame, responsibility, retribution, and justice. Tragedy, too, is inherent in the real possibilities for crime, punishment, and justice, each to be turned into the other. In Marx capital, money, and labor, in which historical and social creations are mistaken as natural ones, is also an "inversion and perversion." This inversion is what Marx called the "fetishism of commodities," with his epithets "Monsieur le Capital" and "Madame la Terre" alluding to the alienation of social relations underlying the uses of capital and land.[49]

Analogously, the three histories here caricature one another. In each, an appearance of reason, through practices that one might want to preserve and promote, contains some germ of irrationality against which reason should protect. The inversion appears either as ironic historical repetition, instructive criticism, or a calamity of irrationalism. Even with Stalinist pathology as the "starting point," its inversion is its positive criticism and return to truth-telling health, and this *did* occur historically. Hungarian Stalinism was a perversion of both Marxism and ideas found in Lakatos, and out of these was still built a revolution of truth. The virtuous punishment of Stalinism inheres in the vices of its own criminality. The revolution also failed, sending thousands into exile, several hundred to their death or imprisonment, and a whole nation into the catatonic self-censorship that was Winston Smith's fate in *1984*. There was no proof, as it were, of the falsification of history that could survive without power, no proof that could prevent the Hungarian Revolution from being rewritten for over thirty years as a counterrevolution. The skeptical lesson is one applied to reason and its value. It cannot be said a priori just where mendacity begins in historical reconstruction, or how to distinguish corrupt falsifications of history by elite writers from the reconstructed pedagogical histories of Lakatos and Hegel, or how to distinguish authentic philosophical self-criticism from public abasement.

For another perspective on Lakatos's verkehrte Welt, or the three worlds depicted above, several of the dramatis personae of the Hungarian Revolution can be compared in terms of how their heroism, compromises, or betrayals played out in their lives. There were certainly some bona fide heroes, including Déry and Nagy, both Communists, anti-Stalinists, and leaders of the revolution, with Déry himself speaking to the crowds gathered at the poet Petőfi's statue on the first day of the revolution, October 23, 1956. At the other extreme are criminals such as Rákosi or Gábor Péter, head of the ÁVH. For these contrasting heroes and villains, their lives relevant to 1956 are marked by a largely unbroken series of either moral or immoral acts. Déry and Nagy, that is, are almost consistently good, and Rákosi and Péter almost consistently evil. But for some others involved in the Hungarian Revolution, including Lakatos, the assessment is complicated by how their actions changed over time and their relative complicity in Rákosi's regime.

Toward the left might be, say, the two chroniclers of the writers' movement, Tamás Aczél and Tibor Méray, both well-known authors who were compromised by their elite status, promoted Stalinism, and then turned against it never to look back.[50] Far to their right, and as close as possible to Rákosi, is Révai. Révai futilely tried to maintain a cynical belief in his own high literary standards while enforcing Moscow's clumsy cultural protocols in place of the Hungarian literature and language he knew as well, and loved as much, as anyone. Révai's compromise was chronic and unfailing. As another case, Rajk was the ultimate victim of immorality, being not a positive hero but a perfect victim of political fabrication. Rajk's life was cut short by betrayal and so compromise would not be a major issue for him. There is also Lakatos, who even after the horrors of the Recsk labor camp, stayed on Rákosi's side until the groundswell of support took over Hungary in later 1955 and then at the Petőfi Circle debates later. But no doubt the greatest figure of compromise and betrayal is Kádár. As Rajk's friend, he convinced him to confess to false accusations, then was accused himself of conspiracy and imprisoned during the Ice Age, later to be released after Stalin's death and Nagy's ascension to prime minister. One might have hoped that Kádár could have transcended his power of betrayal, but alas, he was to turn against Hungary during the last days of the revolution to become a Soviet puppet and the manager of a postrevolutionary asylum. Kádár's lies made his own survival possible at the expense of his nation. Nagy refused to self-criticize, did not lie, and did not survive. Déry did not lie, and so was imprisoned and barely survived. Rajk did not lie but agreed to tell lies about himself, and thus did not survive. Aczél and Méray, and many others including Lakatos, survived by escaping with their lives after deciding not to lie.

These ideal types are intended to help calibrate not Lakatos exactly but his philosophy and the strange histories with which it conspires. His philosophy is simultaneously one of historical reason, compromise, and betrayal. There is one more political-cum-intellectual case, the most important one, whose dissembling and survival is relevant to gaining closure on Lakatos's ideas—that of Lukács. No further information about Lukács's personal or political fate, nor Lakatos's, is needed here. Rather, it is necessary to describe one of Lukács's largest intellectual projects between World War II and 1956: his book *The Destruction of Reason.*

The Destruction of Reason is Lukács's eight-hundred-page history of what he calls irrationalism in nineteenth- and twentieth-century European philosophy, primarily of Germany but also including the United States. Lukács surveys varieties of existentialism, phenomenology, pragmatism, vitalism, and *Lebensphilosophie,* the latter culminating in its contributions to National Socialism and fascism. A principal thesis of the book is that irrational philosophies of the kind Lukács describes provided ideological support for the despotisms climaxing in World War II. In part, the book reflects an impending postwar fear in Hungary of looming international crisis and the repetition of wartime calamities from which Hungary was barely recovering. Lukács characterizes irrationalist philosophies as antidialectical, meaning neither Hegelian nor Marxist; in addition, they substitute for a broad historical rationalism some organizing principle based on, for example, an ineffable existential condition, a subjective experience of internally lived phenomena, or types of pragmatism promoting contextual solutions to political, social, and intellectual problems. These are stereotypes to be sure, but Lukács's position is that existential, phenomenological, and pragmatic worldviews are reactionary in their rejection of history as a primary philosophical category.

This rejection of history is coeval with such philosophies' rejection of Marxism.[51] Lukács considers as irrational any philosophy favoring nonarticulated intuition or experience, and any tendency to invoke atemporal conceptions of consciousness detached from some substantive historical context. The affiliated diminution or tacit rejection of discursive practices generally is also symptomatic of irrationalism in this sense. Examples for Lukács include the philosophies of Søren Kierkegaard, William James, Edmund Husserl, Karl Jaspers, Henri Bergson, Martin Heidegger, and even twentieth-century formal semantics à la Rudolf Carnap. While not mechanically leading to any particular kind of society, these philosophies provide a useful ideological prop for fascist tendencies. So, for instance, "philosophical representatives" of National Socialism, such as Arthur de Gobineau, used the "whole irrationalist development of vitalism in the imperialist period, above all in its final phase . . . to build ideological bridges between Hitlerian propaganda, which could never stoop too low, and the German intelligentsia raised on vitalism," with the SA or SS symbolizing a "vital life-form" in contrast to a fossilized and moribund bourgeois world (*DR* 532). Among the "determining factors" of irrationalism is what Lukács calls a "philosophy of comfort," meaning the suggestion that total freedom can be made possible as an internal, lived experience, while all that is provided is an illusion of personal autonomy, and of moral and intellectual superiority. At the same time, these attitudes are "continually linked with the reactionary bourgeoisie in their real dealings and renders them absolutely subservient to it" (*DR* 22). Rights are "confirmed" through personal intuition, whether in Wilhelm Dilthey, Max Scheler, or Bergson, all of whom are effectively antiscientific and antiobjective.

While Lukács identifies a variety of ways in which antihistorical or ahistorical thinking has been promoted in contemporary philosophy, his broadbrush denunciations and tenuous historical theses struck many as untenable. Theodor Adorno dubs this book "The Destruction of Lukács' Reason"; George Lichtheim, in his study *Georg Lukács,* sarcastically labels it a "contribution" to the "literature" of the "age of the personality cult"; and George Steiner, himself a great admirer of Lukács, in his *Language and Silence* calls

The Destruction of Reason outright "mendacious."[52] These judgments are correct in their way. The book is full of the hack writing typical of the Stalin era, down to the obligatory endorsement of Lysenko genetics, and thus offers abundant evidence for Lukács's intellectual compromises. "What makes *The Destruction of Reason* an essentially Stalinist work," writes Leszek Kolakowski,

> is not, of course, the fact that it seeks the origins of Nazism in German philosophy. . . . The typically Stalinist feature of Lukács' work is the contention that since Marxism came on the scene, all non-Marxist philosophy has been reactionary and irrationalistic. . . . [T]he whole of German philosophical culture outside Marxism is condemned as an intellectual apparatus preparing the way for Hitler's assumption of power in 1933.[53]

That's a fair depiction of the content of Lukács's book. But there is something wrong with the facile use of the label "Stalinist." The book's preface and epilogue are dated November 1952 and January 1953, respectively, both from Budapest. This work, therefore, is the most substantial piece of philosophy to emerge from Stalinist Hungary and during the coldest years of the Ice Age. While Lukács stayed in the background of the writers' movement, as mentioned earlier, he was clearly recognized in Hungary as an independent and courageous Marxist from multiple perspectives: his partisan theory of literature; his battles with Révai during the Rajk affair; his defense of Déry; his general popularity among young intellectuals and critical influence on writers; his reputation for high intellectual standards; his rehabilitation at the Petőfi Circle debates; and of course, his titular role as minister of education for Nagy. Just what should one make of the Hungarian reception to Lukács vis-à-vis appraisals from the West? Who is here being ahistorical? There has to be some other interpretation of *The Destruction of Reason* that reconciles its arguments with the postwar view of Lukács in Hungary.

My suggestion is to focus less on the specifics of Lukács's histories—say, vitalism, phenomenology, existentialism, or pragmatism—than on the book's basic premise, which is mostly neglected. *The Destruction of Reason* uses several primary theses to spin off its perverse history of German philosophy, and these theses also turn out to be directly relevant to Lakatos's work. First is the modern origin of irrationalism in Schelling's esoteric romanticism as attacked by Hegel in the preface to the *Phenomenology.* We earlier saw the relevance of Hegel's attack on Schelling in the *Phenomenology* to the pedagogical philosophy of *Proofs.* That confrontation between history as education and nondiscursive intuition is repeated and articulated at length in *The Destruction of Reason.* Schelling, once again in Lukács's reconstruction, is a dangerous elitist, "acquiring a lasting significance for the development of irrationalism: *epistemological aristocraticism*" (DR 147). Schelling's view of dialectic, according to Lukács, was

> as a mystery, enigmatic and a godsend; talent for dialectical thinking is in that sense eminently comparable with the talent of the artist or genius. Like any gift of the mind it is rare and not something which can be acquired. . . . But the methodology of the *Phenomenology* was directed with great polemical force against that view and proclaimed the universal accessibility, in principle, of dialectics. Indeed without doing violence to Hegel one may say that it was one of the book's chief aims to illustrate the learnable nature of the dialectic, and that one of its purposes as a whole was to guide conventional thinking to dialectics step by step. (DR 574)

It is truly strange how this positive and unobjectionable idea should be neglected in assessments of *The Destruction of Reason.* In fact, Lukács's perspective is simply an update of his 1930s' views and *The Young Hegel.* The update, for example, includes existential authenticity or Aryan racial theory as illustrative of the idea that aristocratic epistemologies divide human beings into two classes: elites in possession of truth and those lacking it. The irrationalisms chronicled in *The Destruction of Reason* are instances of Lukács's conception of elites and their specialized knowledge deriving from the same interpretation of Hegel he held for decades. It is also this pedagogical Hegel that is so useful in reading Lakatos.

Now what about the Stalinism of *The Destruction of Reason?* If aristocratic epistemology is one of the book's underlying themes, the best examples are Soviet Russia during Lukács's years of exile, when he formulated these ideas in *The Young Hegel,* or Rákosi's impeccable performance during the Ice Age. Even Lakatos in his lectures, Paul Feyerabend noted in his obituary for the *British Journal for the Philosophy of Science,* called those elitist theories of knowledge "Stalinist" that were informed by the elites' sure intuition, or *Fingerspitzengefühl.*[54] Regardless of Lukács's intentions, there are no better instances of aristocratic epistemology realized directly in a political regime than those to which Lukács was a direct witness. His experience is all the more remarkable because of what Lukács sees as the fatal turning point in nineteenth-century philosophy. Hungarian Stalinism was not just a dictatorship or totalitarian state. It was a Hegelian inverted world led by elites in which key ideas of Hegelian-Marxism were turned into irrational caricatures. Now one can take the view that this just is all irrelevant to Lukács's intentions, and that he had other notions of the problems of Stalinism. Maybe so. But it is also possible to read Lukács's discussions of elitism, couched in the history of nineteenth- and twentieth-century German philosophy, as covert criticisms of Stalinism's renunciation of exoteric rationality through Party authority; the hierarchical structure of the personality cult; the denial of free and uncoerced speech; and social divisions into bourgeois and working class, or non-Party and Party members.

Lukács's text in this way, in its outlandish and overwritten style, simultaneously epitomizes the central issues of the Hungarian Revolution and implicitly criticizes Hungary's own destruction of reason. Not to compare Hungary's Ice Age to the irrationalist models of the book is willfully naive or positivistic. Lukács, for his part, when asked directly in an autobiographical interview in the late 1960s about *The Destruction of Reason* and Stalinism as a form of irrationalism, thoroughly evaded the question.[55] Self-censorship under Kádár was still endemic in Hungary at that time, and the revolution was still officially a counterrevolution, so it is not clear how effective a comparison of Rákosi and irrationalism would have been. And if the allusions are covert, then there *is* no definitive means of saying it was not intended; the question of authentic meaning is here effectively undecidable. Certainly, burying an acute philosophical description of a tyrannical regime among hundreds of pages of dross is an effective covert strategy, notwithstanding how few Hungarian readers circa 1954 managed to wade through it all. But objectively speaking, a perfectly accurate epistemological description of Stalinism and its critique is hidden in plain sight. This makes the book, whatever its shortcomings and regardless of whether this obscure comparison was remarked by anyone, a quintessential 1956 document, since the centrality of covertness and the changing roles of bad Stalinists into heroic revolutionaries is evident. The truth of *The Destruction of Reason* lay in its expression of all the contradictions animating Hungarian Stalinism.

Lukács's intentions are still to be avoided as much as possible. As such, beyond the observation that *The Destruction of Reason* aptly describes Hungary's personality cult and the Stalinist rejection of Hegelian-Marxism, a second principal theme emerges: Lukács's notion that "there is no such thing as an 'innocent' philosophy" (*DR* 5) vis-à-vis the means by which irrationalist philosophies become "affiliated" with fascism. Regardless of the slanted interpretations Lukács provides, the lesson he draws is that all philosophies are potentially dangerous instruments of coercion when placed in the right political and social context. Nothing immunizes any of them from being turned into wicked ideological tools. Again, the details of how specific ideas led to various political developments is often tenuous. But Lukács is more concerned with the future than the past. The "mere existence of the connecting links we are outlining must be a serious *discite moniti* ['Having been warned, learn!'] for every honest Western thinker. It shows that the possibility of a fascist, aggressively reactionary ideology is objectively contained in every philosophical stirring of irrationalism" (DR 32).

In a 1948 essay titled "On the Responsibility of Intellectuals," Lukács puts it more concretely with a stunning metaphor:

> Finally Hitler took from the reactionary content of these *Salon and Kaffehaus* conversations, university lectures and essays what was usable for his street demagoguery. One cannot find in Hitler one word which had not already been stated by Nietzsche or Bergson, Spengler or Ortega y Gasset ("on a high level"). The so-called opposition of individuals, seen historically, is irrelevant. What significance can a lame, half-hearted protest from Spengler or [Stephan] George have against such a world fire, when their own cigarettes were involved in igniting it?[56]

The belief that philosophical ideas are insulated from political guile or are immune from abuse is a great mistake, according to Lukács, and the corrective is historical reason. One must, as Lukács says, "side with or against reason" (DR 5), which itself is never politically neutral. The "main controversy," philosophically speaking, "between progress and reaction is over the further evolution or destruction of reason" (DR 32).

> Each individual and nation should try and learn something for their own good . . . and this responsibility exists in a particularly acute form for philosophers, whose duty it should be to supervise the existence and evolution of reason in proportion to their concrete share in social developments. . . . So far, not every country has seen the realization of Mephistopheles' lines about the desperate Faust: "Only look down on reason and science / The highest faculties of humankind / . . . And then I have you trapped." (DR 91)

Now the obvious reflexive question is whether Hegel and Marx created innocent philosophies? If *any* philosophies are to be tested for innocence, they should be Hegel and Marx's, especially when the idea of innocence originates with a founder of Western Marxism, a master of Marxist self-criticism, and a victim of the very aristocratic epistemology denounced in

The Destruction of Reason. Not to test Lukács by his own yardstick is to turn a blind eye to the topsy-turvy role of Hegelian-Marxism in Hungary. The answer is by this stage obvious: key ideas of Hegelian-Marxism were themselves turned into demagogic tools. Hegel and Marx's ideas are not therefore "guilty," meaning that these changes were neither automatic nor necessary. It took a much greater social and political history to create Stalinism, not just the crafty reformulation of ideas. Nevertheless, the ideas played a role, and a powerful one at that. They are not innocent of the grotesque transformations of self-criticism, reconstructed history, and Party leadership that occurred in Lukács and Lakatos's Hungary. Certainly, these ideas are as implicated in the development of Stalinism as any of the other philosophies chronicled by Lukács vis-à-vis Hitler and the rise of fascism.

Again, Lukács never even hinted at the interpretation of Hungarian life before 1956 as a perversion of his own Hegelian-Marxism, but that still is what he depicted and what actually happened in Hungary. Such may be the closest Lukács comes to the complicity of his own philosophical tradition in one of the great twentieth-century tragedies. While it might be preferable if this reading were more "explicit," that is not consistent with Lukács's real survival mode. Like other Hungarian intellectuals, he was a master of tact, ambiguity, and concealed reproach. He sometimes suggested that he was a kind of Trojan horse under Stalinism, and his works, as Kolakowski explains, "are full of palinodes, retractions, withdrawals of retractions, and reinterpretations of earlier writings."[57] *Hic Rhodus, hic saltus:* the Hungarian Revolution was for truth, but was created through cunning, and thus, *The Destruction of Reason* appears as an exquisite shape of compromised consciousness and reason. It is historicist rationality, Stalinist perversion, and cunning political criticism all in one. If no philosophy is innocent, then so too every philosophy is suspect, and artful concealment is one's only protection.

Lukács was no doubt compromised along with many other admirable figures of the Hungarian Revolution. Yet after about 1950, he is rather a philosopher of compromise and survival. Not everyone could be a Nagy or Déry, and it might just be rational to lie for you and your ideas to survive. One might end up expressing that too, implicitly, in how one's philosophy is created, even if not the intent, especially when the category of intention is rather battered in a paranoid world of Orwellian falsification. That is how one might survive the fall from innocence. *The Destruction of Reason is* a compromised work, as Adorno and others have contended. But "no philosophy is innocent" correctly expresses the truth for Hungary that the philosophy of Hegelian-Marxism in its perverse form was at the heart of Hungary's woes.

And it was nonetheless still the Communist writers who initiated the revolution from within the Party and, for thirteen days, managed the triumph of good over evil. *The Destruction of Reason* does grasp in philosophical thought, just because it does not transcend, the real contradictions of Hungary's most recent heroic and tragic era.

Lukács, at least until his return to Hungary, should also not be underestimated in his ability to embrace militancy when needed. In *History and Class Consciousness,* he heaped scorn on "bourgeois civil liberties" and eagerly promoted the enforcement of totalitarian Party discipline.[58] During the Hungarian Communist Revolution of 1919, Lukács at one point joined a military offensive along with many other political commissars. As Lukács put it himself,

> The defense of Tiszafüred had been grossly mismanaged because the Budapest Red Army units ran away without firing a shot . . . and Tiszafüred fell. I set about restoring order as energetically as I could. That is to say, when we crossed the river to Poroszló, I set up a court-martial and had eight men belonging to the battalion that had run away in panic shot in the market-place. By these means I more or less managed to restore order.[59]

Does that past, any more than Lakatos's, affect Lukács's intentions as they relate to *The Destruction of Reason?* No: it means that Lukács's own life is also exemplary of the betrayals and compromises that came to a head in postwar Hungary.

For another perspective, Lukács has long been recognized as the model for the Jesuit Naphta in Thomas Mann's novel *The Magic Mountain.* Naphta rooms in a house owned by the Lukaçek family, and is described by Mann as physically small and unattractive like Lukács. He is a brilliant rationalist with a secret commitment to world socialism. At least from the rarefied safety of a sanitarium high in the Swiss Alps, Naphta supports terror and even torture as rational means of necessary hierarchical control. Naphta's essence is not so much that of fanatical rationalist-socialist; rather, it is his uncanny ability to simultaneously promote and attack what seem to be totally antithetical doctrines—say, Christian love and terror, or free individuality and rigid socialism. His moral world is the Manichean mix of good and evil as expressed in our epigraph to part 3. Mann's acute portrayal of Lukács from the 1920s requires some qualification, for Lukács under Hungarian Stalinism had by that time few illusions about the failures of his government and its political system. For example, only days before Soviet tanks invaded

Budapest, Lukács remarked to a reporter from *Nowa Kultura* that "The new Party will not be able to expect rapid success—Communism in Hungary has been totally disgraced. . . . In free elections the Communists will obtain five percent of the vote, ten percent at the most. . . . [The Party] will be an intellectual center, and after some years or some decades from now, who knows."[60] Nonetheless, *The Destruction of Reason* is for us the gestalt of a contradictory and extremist Naphta: the book implicitly acknowledges Communism's disaster while maintaining Marxism as the only true and rational philosophy, prostrating to Stalinist orthodoxy while subverting it subtly from within. The elderly Naphta expressed how in Hungary good and evil had become one with his cherished historical reason.

Lakatos, or rather his essays, interestingly leads to *The Destruction of Reason* through a posthumously published polemical review of Stephen Toulmin's *Human Understanding,* titled "Understanding Toulmin" (*MSE* 224–41).[61] One of the essay's primary themes is epistemological elitism, with an emphasis on epistemologies accounting for knowledge through private and nondiscursive means. Lakatos repeats Lukács's claim that pragmatist philosophies are essentially elitist and an intellectual source of fascism in English-language philosophy. That idea about pragmatism, which Lakatos can even use as the philosophical heritage of his contemporary Quine, is the only Anglo-American variety of irrationalism discussed by Lukács, whom Lakatos mentions in passing without citing him as a source. Lakatos also makes some fairly strained interpretations to fit Toulmin's Wittgensteinian ideas into the irrationalist camp. The later Wittgenstein of the *Philosophical Investigations,* for example, is unorthodoxly classified as a pragmatist, and hence an elitist, which is certainly an odd characterization of the philosopher who created the so-called private language argument for the necessarily public and intersubjective nature of linguistic meaning. It is also not as if Lakatos had not read Wittgenstein, for his footnotes show that he has studied him well. Toulmin, Michael Polányi, and Thomas Kuhn fare not much better here, and the shrill and dogmatic essay provided some puzzlement for many of Lakatos's colleagues. During the early 1970s, sociological or quasi-anthropological accounts of science were still novel enough to represent a radical approach to problems of knowledge, but for Lakatos, they represented an irrationalist threat to discursive reason and science, of which Lukács warned in the 1950s. Lakatos wrote as if he was hearing Mephisto: "Only look down on reason and science . . . And then I have you trapped."

Once one knows the ideas of *The Destruction of Reason,* Lakatos's polemic and its reactionary stance appear as almost a set piece. Lakatos has done nothing more than to re-create in small, and in the words of anglophone philosophy, one central idea of his old teacher's book. Not only that, Lakatos's rhetoric managed to evoke the same response from his audience that he was over the top, that his own dogmatic rationalism had gone amuck. There is even a neat consistency between Lukács and Lakatos's appeals, in response to the threats of irrationalism, to respect science and reason. What should be made of that? Do we have an imitation bordering on parody of one of the strangest books in the Hegelian tradition, itself knotted up in historical and cultural paradoxes? A hyperbolic covert reference to a hyperbolic covert criticism of Hungarian Stalinism whose positive antidote, the Bildungsphilosophie of the *Phenomenology,* is also a model for Lakatos's *Proofs?* A covert reference to a former teacher whose skeptical self-application of Marxism is repeated as the tool to create Lakatos's historiographical theory? It would be convenient here to use the "external" history of Lakatos's life to help sort this out, but where does that lead?

Take, for example, one of Lakatos's extreme reversals—compared to his dark days in Hungary—which he underwent as an illustrious professor of philosophy at the London School of Economics. During the student movement of the late 1960s, Lakatos was well known as an academic and political conservative as well as a supporter of the Vietnam War. He taunted his enemies by wearing an "Agnew for Vice President" button, though here too there may have been an element of maintaining a joking relationship with his political antagonists. At the time of the student shutdown of the LSE, Lakatos wrote an open letter to the LSE director imploring him and other university leaders to protect academic autonomy, and reject student demands for stronger participation in academic decision making.[62] Over a thousand copies of the letter were reproduced and it was influential in the return to power by the administration following the student strike. Lakatos was a hero for eloquently articulating a successful defense of academic independence. Having survived the purgatory of a genuine political revolution in Hungary, himself an ex-Marxist and a promoter of progressive philosophical views, Lakatos's credentials were also politically correct. What nobody knew was that the letter to the director of the LSE was also the work of the charismatic Communist who two decades earlier had brought down Eötvös College in Budapest with his brilliant oratory and writing. Nobody also knew Lakatos had been the fanatical leader of a Communist cell and that he apparently orchestrated the death of an innocent woman.

Was Lakatos making right for his previous wrongs? I hope so, though he still erred in supporting an unconscionable war of a nation not his own; he was also not even a British subject, it turns out, since his dodgy Hungarian background apparently kept him from receiving citizenship. Can one be so cynical as to think Lakatos might *not* be making amends for his past, that his actions at the LSE were anything less than honest? Yes, although it is not a question of cynicism but skepticism—skepticism about the actions of someone who earlier in his life was a dangerous thug with something like a criminal record, and who consistently displayed a pattern of dissemblance and cunning across the decades.

The ambiguities of Lakatos's life to the end do not remove the contradictions of his philosophical work, and may even be taken to reinforce them. As much as we might like to "resolve" Lakatos's philosophy toward goodness, progress, and hope, that would be a naive, unjustified, and even hypocritical gesture of faith missing the entire thrust of Lakatos's work and its histories. The external history of Lakatos's life, in other words, if invoked, can be just as well taken to reinforce the moral ambiguities of reason found in his philosophy as it can be used to turn them into some comforting philosophical lesson. Protagoras is the philosopher who is said to have first proposed the skeptical possibility of isostheneia, in the form that equally strong arguments can always be mobilized for both sides of a debate. Lakatos's consistent Hegelian commitments make that just as possible for his life as for his ideas. So if through his life you want to evade the unsettled and morally challenging endings to which his ideas lead, then you will not succeed.

The significance of the "Understanding Toulmin" essay, written so close to the end of Lakatos's life, should not be overemphasized. But through it, as with *The Destruction of Reason,* the compromises of reason and its paradoxes can be witnessed. We began this book with the problem of understanding Lakatos and the Hegelian histories with which he appeared to be implicated. The original problem was that of Lakatos's ingenious covert attack on his mentor Popper. Did we solve the problem we set out to solve? Did we learn what we set out to learn? We are left with the task of attaining closure on the problem of Imre Lakatos and the multiple layers of his intellectual past.

I take Lakatos's weird repetition of Lukács's polemic in *The Destruction of Reason* as an occasion to borrow Lukács's lesson that no philosophy is innocent also as Lakatos's lesson as a cross-cultural educator, a Hegelian Volkserzieher. Here I mean the lessons of the three worlds—English-language philosophy, Hegelian-Marxism, and Hungary—and the evident affinities and caricatured relations among them. Lakatos created a powerful historical rationalism. Had you been a Hungarian from the end of World War II to 1956, it would have been a remarkably powerful set of ideas to have in your head, even if you could not utter a word of it to anyone. Lakatos's philosophy is a rationalism not least because it could help keep one sane in an insane world. But Lakatos's philosophy is not all powerful. More than that, a horrible incarnation of it once existed that it has *some* power to protect against, but without any guarantees.

To be sure, it is remarkable how Lakatos's work answers just about all the problems set for a historical philosophy of science coming out of the history of Hegelian-Marxism and the intellectual politics of Stalinist Hungary. Nonetheless, the first lesson of innocence is skepticism about the moral content of historical reason and the ability to protect reason from abuse. The potential pessimism of that lesson has a corollary, though, which is that the promoters of reason can still be sly and cunning, and use reason to survive and prosper. Not too many noticed, but that is what Lukács in the early 1950s implied *could* happen, and is what *did* happen in Hungary in 1956. Lakatos wanted criticism to be open, discursive, and exoteric, as did Popper. Lakatos offered excellent means for all that to be accomplished, as well as the means for diagnosing its failures in mathematics and science. But that does not imply that all good criticism is out in the open. Lakatos's Hegelian criticism of Popper was covert, as was Marx's use of Hegel against the political economists, as was that of Hungarian writers who turned against an abhorrent irrationalism. Lakatos's covertness, like his teacher Lukács's, is a part of his philosophical rationalism and evidence that it is not innocent. That lack of innocence is sometimes a force for good, sometimes a force for evil, with considerable permeability at the boundaries. Lakatos said in an aside during a debate in England that he once intended to be Lukács's successor. Much of Lakatos's philosophy supports that, including the shared interest in Bildungsphilosophie, the antielitism, and the use of reflexive tropes of self-application; now, too, the polemical irruption against irrationalism and the clever philosophical attacks from within.

There also are differences between how the failure of innocence appears in Lakatos and Lukács. First, while Lukács's appeal to reason is a deus ex machina, Lakatos has a detailed account of scientific criticism and the historiography of science, even one that can be used to analyze the scientific status of Marxian political economy. Similarly, Lakatos's Bildungsphilosophie in *Proofs* provides a historical and exoteric portrait of mathematical reason with a substantive identification of Lukács's aristocratic epistemology in the ideology of Euclidean method.

Second, more philosophical arguments for Lukács's claims about innocence exist in the mitigated skepticism appearing throughout Lakatos's work. The good and evil sides of historiography, for example, are represented through Lakatos's methods of rational reconstruction and its Stalinist-Orwellian twin. In Lakatos's account, the relation between reconstructed history and so-called actual history is that of theory ladenness and the inexorable dependence of all historical reconstructions on some value-laden historiographical technique. Not to oversimplify, but at bottom, that is just a skeptical claim about the relational status of historiographical interpretations vis-à-vis their historical objects. "No philosophy is innocent" can be taken as shorthand for the skeptical possibility that Lakatos's historical rationalism can be deployed with opposite moral polarities.

A third difference with Lukács is that Lakatos's work comes after Lukács and the Hungarian Revolution. After Lukács: who would have dreamed that Lukács's recovery of Marx's covert use of Hegel could itself become an implicit topic of a historical philosophy appearing through its own covertness? After Hungary: as stated before, Lakatos's work can be seen as a fairly complete response to Stalinism with respect to *exactly* the problems of falsified history, elitism, science, and self-criticism that dragged Hungarian intellectuals into the swamp of the personality cult.

Yet, in spite of these differences, the final bond between Lakatos and Lukács is that both are Naphta's cunning progeny. Both derive, express, and live out good and evil through the power and guises of historical reason. Both are far from innocent in their political histories, and both end up showing that the dissemblance and cunning by which they contributed to and survived all kinds of terrors is part of reason, too. All this can be learned from the problem of Lakatos's concealed Hegelian critique of Popper, the problem with which this book began some afternoons ago.

The world of reason, I conclude, is a world of social and intellectual relations stood on its head. It is an enchanted, perverted, and topsy-turvy world in which Monsieur le Langage and Madame la Histoire do their ghost walking as social characters, and at the same time directly as mere things, as forces partially under our control, but whose power over us can become total. Not only are there no foundations for knowledge, we are constantly subject to the danger of creating foundations for the antiethics of Stalinism, the antiscience of Lysenko genetics, and the antihistories of Muscovite historians. Our glance at Hungarian Stalinism shows that historical writing may be stood on its head, then heroically turned over again, and that historical reason is a body of strategies waiting to be seized. And almost anybody can seize them. "The analogy between political ideologies and scientific theories," Lakatos writes, "is then more far-reaching than is commonly realized" (P&R 49n1). *Caveat lector!*

Notes

Abbreviations for Works Cited

C&GK Imre Lakatos and Alan Musgrave, eds., *Criticism and the Growth of Knowledge* (New York: Cambridge University Press, 1970).

DR Georg Lukács, *The Destruction of Reason,* trans. Peter Palmer (Atlantic Highlands, N.J.: Humanities Press, 1981).

FM Imre Lakatos, "Falsification and the Methodology of Scientific Research Programmes," in *The Methodology of Scientific Research Programmes: Philosophical Papers, Volume I,* ed. John Worrall and Gregory Currie (New York: Cambridge University Press, 1978).

H&CC Georg Lukács, *History and Class Consciousness: Studies in Marxist Dialectics,* trans. Rodney Livingstone (Cambridge, Mass.: MIT Press, 1971).

HS Imre Lakatos, "History of Science and Its Rational Reconstructions," in *The Methodology of Scientific Research Programmes: Philosophical Papers, Volume I,* ed. John Worrall and Gregory Currie (New York: Cambridge University Press, 1978).

LSCD Karl Popper, *The Logic of Scientific Discovery* (New York: Harper, 1959).

MA Colin Howson, ed., *Method and Appraisal in the Physical Sciences: The Critical Background to Science, 1850–1900* (New York: Cambridge University Press, 1976).

MSE Imre Lakatos, *Mathematics, Science, and Epistemology: Philosophical Papers, Volume II,* ed. John Worrall and Gregory Currie (New York: Cambridge University Press, 1978).

MSRP Imre Lakatos, *The Methodology of Scientific Research Programmes: Philosophical Papers, Volume I,* ed. John Worrall and Gregory Currie (New York: Cambridge University Press, 1978).

P&R Imre Lakatos, *Proofs and Refutations: The Logic of Mathematical Discovery,* ed. John Worrall and Elie Zahar (New York: Cambridge University Press, 1976).

Pf G. W. F. Hegel, preface to *Phenomenology of Spirit,* in *Hegel: Texts and Commentary,* trans. Walter Kaufmann (Garden City, N.Y.: Anchor-Doubleday, 1965).

PhS G. W. F. Hegel, *Hegel's Phenomenology of Spirit,* trans. A. V. Miller (New York: Oxford University Press, 1979).

RH G. W. F. Hegel, *Introduction: Reason in History,* in *Lectures on the Philosophy of World History,* trans. H. B. Nisbet (New York: Cambridge University Press, 1975).

RM Tamás Aczél and Tibor Méray, *The Revolt of the Mind: A Case Study of Intellectual Resistance behind the Iron Curtain* (New York: Praeger, 1958).

SSR Thomas Kuhn, *The Structure of Scientific Revolutions,* 2d ed. (Chicago: University of Chicago Press, 1970).

TSV Karl Marx, *Theories of Surplus-Value,* trans. Emil Burns, 3 vols. (Moscow: Progress Publishers, 1963).

Introduction

1 For the 1965 London School of Economics conference on which C&GK is based, Lakatos's battle with Popper, and the intellectual milieu, see Luce Giard, introduction to *Histoire et Methodologie des Sciences: Programmes de recherche et reconstruction rationnelle* (Paris: Presses Universitaires de France, 1994).

2 See Karl Popper, *Objective Knowledge: An Evolutionary Approach* (Oxford: Oxford University Press, 1972), 122.

3 G. W. F. Hegel, *Hegel's Philosophy of Right,* trans. T. M. Knox (New York: Oxford University Press, 1979), 11.

4 "Historicism" here refers to philosophers or philosophies making substantive use of historical or historiographical categories and materials. The term implies nothing about whether those categories are social, institutional, linguistic, material, or some combination of all; nor whether historical processes proceed deterministically, mechanically, or through the conscious intervention of human agency; nor whether historical change is continuous or not; nor anything about the historiographical forms used to reconstruct the past. Kuhn, Lakatos, Feyerabend, Marx, Hegel, and Michel Foucault are historicists in this sense, while W. V. O. Quine, Rudolf Carnap, John Austin, and Ludwig Wittgenstein are not. Popper used historicism to imply the means for historical prediction based on patterns or trends underlying historical evolution; see, for example, Karl Popper, *The Open Society and Its Enemies,* 4th ed. (Princeton, N.J.: Princeton University Press, 1963), 1: 8.

5 Mark Blaug, "Kuhn versus Lakatos *or* Paradigms versus Research Programs in the History of Economics," in *Method and Appraisal in Economics,* ed. Spiro Latsis (New York: Cambridge University Press, 1976), 150. Blaug is commenting on Lakatos's "tendency to make vital points in footnotes, to proliferate labels for different intellectual positions, and to refer back and forth to his own writings" (ibid.). Feyerabend seems to have grasped the essential function of Lakatos's footnotes. His *Against Method* was originally to have been the first half of a joint work on rationalism whose second half would have been Lakatos's defense answering Feyerabend's attack; the book is subtitled *Outline of an Anarchist* Theory of Knowledge* with the asterisk directing the reader, on the title page no less, to a footnote within the text on the meaning of anarchism. *Against Method* is also dedicated "To Imre Lakatos/Friend, and fellow-anarchist." Friedrich Engels, too, mentioned "Marx's art of quotation which is so little understood." These quota-

tions occur mostly in the footnotes and form "a running commentary to the text, a commentary borrowed from the history of economic science" (preface to *Capital,* by Karl Marx, 3d ed. [New York: International Publishers, 1967], 1: 24–25)—that is, much like the narrative technique used throughout *P&R*.

Lakatos also adapts Hegel's technique from *Phs* of using perhaps unnoted, partially modified quotations—for example, Hegel's closing lines adapted with modification from Friedrich Schiller. Lukács remarks on his own frequent quotations in *H&CC*, asserting that "every quotation is an interpretation" (xliii), but quotations also served the prosaic function in Lukács's work of providing textual materials to readers for whom they might otherwise be unavailable: "Hungary got philosophical writing and teaching characterized by an incessant citing of Marxist-Leninist classics [that is, 'citationism']. . . . Many students of philosophy [in the early 1960s] learned all they knew of non-Marxist philosophers from the extensive quotations from their works in Lukács' writings" (Thomas Szendry, "Philosophy in Hungary before and after 1956: From Staunch Orthodoxy to Limited Revisionism," in *The First War between Socialist States: The Hungarian Revolution of 1956 and Its Impact,* ed. Béla Király et al. [New York: Brooklyn College Press, 1984], 223–24). The many quotations in this chapter are intended to make the reader aware of the layers of embedded concepts, stylistic traits, and methods through which ideas are transformed and appear in Lakatos's work. In literature, extensive disguised quotation was used by Thomas Mann in *Doctor Faustus,* a technique Mann called "montage." Mann's motivation was that he wanted to show in his own work the modernist crisis of being condemned to certain forms of repetitious "quotation," whether in musical composition, the subject of the book, or literature. One of Mann's sources for this idea was Theodor Adorno, who appears in the novel as one incarnation of the devil. See Gunilla Bergsten, *Thomas Mann's "Doctor Faustus": The Sources and Structure of the Novel,* trans. Krishna Winston (Chicago: University of Chicago Press, 1969).

6 There is no reason for the scare quotes around "dialectically."

7 Quoted in the editor's note. Apparently, Hegel was deleted from later versions; see *P&R* (xii), where only Pólya and Popper are mentioned.

8 See *P&R*, 144–46. On Hegelian triads, "Fichte introduced into German philosophy the three-step of thesis, antithesis, and synthesis, using these three terms. Schelling took up this terminology; Hegel did not. He never once used these three terms together to designate three stages in an argument or account in any of his books" (Walter Kaufmann, *Hegel: A Reinterpretation* [Garden City, N.Y.: Anchor-Doubleday, 1965], 154). Hegel even criticizes the triad as "picture-thinking," the wrong type of *Vorstellung,* akin to the difference between religion and philosophy (*Phs*, 469; also *Hegel's Science of Logic,* trans. A. V. Miller [New York: Humanities Press, 1976], 846). Popper often attributes the triad to Hegel and generally mocks his ideas (for instance, "What Is Dialectic?" in *Conjectures and Refutations: The Growth of Scientific Knowledge* [New York: Harper Torchbooks, 1968], 313–14). Kaufmann describes Popper's version of Hegel as depending on the selective "quilt quotation" of abridged passages taken out of context; on Popper's Hegelian infelicities, see Walter Kaufmann, *From Shakespeare to Existentialism* (Garden City, N.Y.: Anchor-Doubleday, 1959), chapters 7 and 8. It is then wickedly ironic to see

that Lakatos, as the most zealous disciple of the author of the influential anti-Hegelian *Open Society and Its Enemies,* was the Hegelian Trojan horse, the avenger who stole the citadel from within; we are witness here to quite an object lesson in the meaning of critique. Nonetheless, key themes of *The Open Society* reappear in Lakatos's work. A principal idea of *The Open Society* is that totalitarianism is engendered by a belief in historical determinism (and "prophecy") and some type of elitist or authoritarian conception of knowledge or society, both of which Popper finds in Plato, Hegel, and Marx. Lakatos is a historicist but no determinist, while he retains a role for elites in judging scientific progress. At the same time, his native Hungary was for many years a totalitarian caricature of Marxism and a closed society of the kind Popper warned against. My view of Lakatos in chapter 12 is that while he substantially corrects Popper's ideas about history and criticism, his own historicism (principally his historiography) preserves ideas that in a perverse form were at the center of Hungarian Stalinism. In this way, the spirit but not the letter of *The Open Society* is implicitly maintained in Lakatos's work.

9 Marx Wartofsky, "The Relation between Philosophy of Science and History of Science," in *Essays in Memory of Imre Lakatos,* ed. Robert Cohen et al. (Boston: D. Reidel, 1976), 726.

10 Paul Feyerabend, "Against Method," in *Minnesota Studies in the Philosophy of Science IV,* ed. M. Radner and S. Winokur (Minneapolis: University of Minnesota Press, 1970), 100 n. 27; see *Against Method* (London: New Left Books, 1975), 17–130, 100 n. 27; compare *Against Method,* 200, for the "Trojan horse," with that description also used for scientific observational terms in Galileo (75).

11 Ian Hacking, "Imre Lakatos' Philosophy of Science," *British Journal for the Philosophy of Science* 30 (1979): 383. Though Hacking does not specify the Hungarian element, it is Lukács's work that provides exactly the right Hegelian content.

12 "Unaware of the existence of the *Economic-Philosophical Manuscripts,* Lukács none the less succeeded in reading the Hegelian issue of alienation back into Marx's later writings, and thus established the importance of alienation in Marx's theory. This was an outstanding intellectual feat, and the subsequent discovery of the *Manuscripts* confirmed most of Lukács' insights" (Shlomo Avineri, *The Social and Political Thought of Karl Marx* [Cambridge, U.K.: Cambridge University Press, 1971], 96). The famous error of Lukács's interpretation was his confounding reification, objectification, and alienation, leading also to a misplaced emphasis on a cultural-critical versus social-theoretical and economistic Marx.

13 The two sentences in the text involving the "dead dog" and so forth from the above paragraph with "Marx" instead of "Lakatos" are from the 1922 preface to *H&CC* (xliv); they end with, "and that in the interests of 'scientific precision' all traces of it [that is, Hegel] should be eradicated systematically from the method of historical materialism." (The original dead dog discussion about Hegel is from Marx's afterword to the second German edition of *Capital,* 1: 19.) For slight discussions of Hegel and Marx in Lakatos, see *P&R*, 146, and *MSE*, 125.

14 Quoted in David Krell, "The Mole: Philosophic Borrowings in Kant, Hegel, and Nietzsche," in *Why Nietzsche Now?* edited by Daniel O'Hara (Bloomington: Indiana University Press, 1985), 169. Marx borrowed the mole metaphor from Shakespeare's *Hamlet* and used it to refer to submerged revolutionary forces.

15 See H. S. Harris, *Hegel's Development: Towards the Sunlight, 1770–1801* (New York: Oxford University Press, 1972), 159, 163–65.
16 William Bartley III, "On Imre Lakatos," in *Essays in Memory of Imre Lakatos,* ed. Robert Cohen et al. (Boston: D. Reidel, 1976), 37–38.
17 Feyerabend, *Against Method,* 200; and Paul Feyerabend, "Imre Lakatos," *British Journal for the Philosophy of Science,* 26 (1975): 1.
18 Bartley, "On Imre Lakatos," 37–38.
19 Feyerabend, "Imre Lakatos," 2.
20 Feyerabend, "Against Method," 200; see also 115 n. 27, 125 n. 91.
21 Lakatos quoting Popper in *P&R*, 105; see also *P&R*, 6.
22 Lukács on Rosa Luxemburg and rational reconstruction in *H&CC*, 34.
23 I do not address Lakatos's work on Cauchy and nonstandard analysis (*MSE*, chapter 3), or his lengthy demolition of Carnap's inductive logic (*MSE*, chapter 8), neither of which I think are central to his philosophy. For a critical discussion of Lakatos's work on Cauchy and nonstandard analysis, see Teun Koetsier, *Lakatos' Philosophy of Mathematics: An Historical Approach* (Amsterdam: North-Holland, 1991), chapter 3.
24 Hacking, "Imre Lakatos' Philosophy of Science," 387.
25 Quoted in *MSE*, 131.
26 Popper, *Conjectures and Refutations,* vii–ix.
27 Popper, *Objective Knowledge,* 20, 186.
28 Quoted in Georg Lukács, *The Young Hegel: Studies in the Relations between Dialectics and Economics,* trans. Rodney Livingstone (Cambridge, Mass.: MIT Press, 1976), 254. Completed in 1938 while Lukács was exiled in the Soviet Union, *The Young Hegel* contains much of the relevant Hegelian perspective for *P&R*, including *PhS*'s pedagogical goals, its bildungsroman form, and Hegel's antiromanticism. As discussed here in chapters 1 and 12, it is important also that the Stalinist personality cult and its authoritarian terror, which Lukács survived, is aptly characterized through the irrational elitism Lukács described in Hegel's critique of Schelling. Lukács never admitted to the clear parallel with Stalinism.
29 In this process, the key word *falsification* is transformed by Lakatos to include the falsification of Popper's philosophy by the history of science, and then the falsification of history generally. As discussed in chapter 11, the latter idea plays a prominent role in Marx's history of economics, *Theories of Surplus-Value.*
30 Hence, there is a clear connection to Lukács's literary expertise via historiographical techniques transported by Lakatos into the history and philosophy of science and mathematics. Lukács's attention to narrative form suggests how the historiographical methods he describes might be used in different intellectual contexts.
31 The word *dialectic* itself is as tainted as *incommensurability,* of which Feyerabend once remarked that everyone who deals with it ends up with mud on their face. Lukács was said to overuse and abuse the term *dialectic* as much as anyone (see Leszek Kolakowski, *Main Currents of Marxism,* trans. P. S. Falla (New York: Oxford University Press, 1981), 3: 305.
32 The term is from George Steiner's *Extraterritorial: Papers on Literature and the Language Revolution* (New York: Atheneum, 1976).

1 | The Mathematical Present as History

1 Lakatos's Hegelianism refers not to his intentions but the logic, methods, concepts, and implications of his work.

2 The quotations on Lakatos's "absurd historiography" (Nathan Reingold) and "parody" (Gerald Holton) are in Ian Hacking, "Imre Lakatos' Philosophy of Science," *British Journal for the Philosophy of Science* 30 (1979): 383. The comments were made with respect to Lakatos's philosophy of science, but the issue of historical falsification versus reconstruction is the same. See chapter 9 herein for a continued account of Lakatos's historiography, note 25 in chapter 8, and this chapter's quotation of Rudolf Haym, text to note 20.

3 Paul Feyerabend, "Imre Lakatos," *British Journal for the Philosophy of Science* 26 (1975): 14.

4 On Alfred Tarski and Gödel, see *P&R*, 3 n. 3.

5 I use "mathematical logic" to refer generically to mathematical methods in foundational studies, including logic, set theory, proof theory, and even metamathematical techniques such as Gödel numbering. The justification for this imprecision is that *P&R* is oriented around formalization generally and not its specific expression in any single area of logic.

6 On Gödel's theorems, see *P&R*, 2 n. 2; and *MSE*, 37. For much more on Gödel's theorems, see chapter 4 below.

7 On "dominant theories," see *P&R*, 123–25, and chapter 4 below. W. V. O. Quine's holism begins in his "Two Dogmas of Empiricism," in *From a Logical Point of View* (Cambridge, Mass.: Harvard University Press, 1961); on many-valued logic, see Hilary Putnam, "The Logic of Quantum Mechanics," in *Mathematics, Matter, and Method: Philosophical Papers, Volume I* (New York: Cambridge University Press, 1975).

8 On Marx's use of footnotes and historical quotation, see introduction, note 5.

9 This paraphrases Lukács on Marx in *H&CC*, xx.

10 Quoted in Georg Lukács, *The Young Hegel: Studies in the Relations between Dialectics and Economics,* trans. Rodney Livingstone (Cambridge, Mass.: MIT Press, 1976), 468. Engel's *Ludwig Feuerbach* discusses Marx's historiographical style at length and Marx's extraction of a logical-historical kernel out of Hegel. In Marx, these ideas are to be found largely in his *Grundrisse;* see chapter 5, note 9, below, and text. On Engel's biogenetic description and the methodology of *Capital,* see Ronald Meek, *Studies in the Labour Theory of Value,* 2d ed. (London: Lawrence and Wishart, 1973), 320.

11 See Loren Eiseley, *Darwin's Century* (Garden City, N.Y.: Anchor Books, 1962), 95. For the scientific history of biogenetic law, see Stephen J. Gould, *Ontogeny and Phylogeny* (Cambridge, Mass.: Harvard University Press, 1977).

12 Quoted in Hans Reiss, *Goethe's Novels* (New York: St. Martin's Press, 1969), 122. Much of my account of *Wilhelm Meister* draws on Reiss.

13 Quoted in Johann Wolfgang von Goethe, *Wilhelm Meister,* trans. Thomas Carlyle, vol. 2 (New York: E. P. Dutton, 1824), 61. Apropos of Hegel, see also Wilhelm's thought that "I cannot bear to look on people making awkward trials. . . . On this point, I had a continual struggle with the Abbé, who maintains that error

can never be cured except by erring" (ibid., 106). *Wanderjahre* denotes the period in which a German artisan travels about to perfect his or her craft after a *Lehrjahre,* or apprenticeship, and before a mastership begins.

14 Quoted in Reiss, *Goethe's Novels,* 129.

15 Ibid., 134.

16 Ibid., 131; from a July 1796 letter from Schiller to Goethe.

17 See also Marx's "The Eighteenth Brumaire of Louis Bonaparte": "Men make their own history, but they do not make it just as they please; they do not make it under circumstances chosen by themselves, but under circumstances directly found, given and transmitted from the past. The tradition of all the dead generations weighs like a nightmare on the brain of the living" (in *The Marx-Engels Reader,* ed. Robert Tucker, 2d ed. [New York: W. W. Norton, 1978], 595).

18 See Michael Forster, *Hegel's Idea of a Phenomenology of Spirit* (Chicago: University of Chicago Press, 1998), 292 ff., supporting Lukács's account of the historical aspects of *PhS*, and Forster's explication and defense of Hegel's historical sequencing.

19 As Lakatos points out (*P&R*, 139 n. 1), Hegel believed mathematics to be outside the scope of fallibilism, hence implicitly outside history and Bildung, and more generally, impervious to an integration of Kantian *Verstand* (understanding) and Hegelian *Vernunft* (reason). Since the publication of *PhS*, the possibility of a genetic and historical account of mathematical concepts has never been conceived, much less for the same pedagogically and historically motivated reasons as enunciated in Hegel's preface to it. For this reason, the importance of Lakatos's achievement in the history of ideas should not be underrated. As expert a commentator and practitioner of Hegel's philosophy of cultural education as Hans-Georg Gadamer denied in his *Truth and Method* (trans. Joel Weinsheimer and Donald Marshall [New York: Seabury, 1975], 77) that Bildung had any relevance to mathematics. And as expert a scientist, philosopher, and historian of science as Pierre Duhem, in *The Aim and Structure of Physical Theory* (trans. Philip Weiner [Princeton, N.J.: Princeton University Press, 1954], 268–70), made the same negative assertion when he discussed the relevance of the narrative analogue of the biogenetic law to the conceptual foundations of modern science. Hence, Lakatos has carried out a historical critique of mathematical formalism in a manner wholly unsuspected by Hegel, and that would become paradigmatic of nineteenth-century attempts to historicize both human nature and social institutions. As Hegel himself stated in his long discussion of mathematical proofs, "A critique of these demonstrations would be as remarkable as it would be instructive and might both cleanse mathematics of this false finery and show the limitations of mathematics and thus also the necessity of another kind of knowledge" (*Pf*, 68). Another kind of knowledge indeed.

20 Quoted in Lukács, *The Young Hegel,* 467. See this chapter's note 2 above for contemporary reactions to Lakatos's historiography. Lakatos moves well beyond Hegel insofar as he provides pervasive criticisms and historical documentation of a kind glaringly absent from *PhS*.

21 H. S. Harris, *Hegel's Development: Towards the Sunlight, 1770–1801* (New York: Oxford University Press, 1972), 195; Harris is here commenting on Hegel's 1795

"Life of Jesus." The combined interest in historiographical techniques and cultural learning deployed by Hegel's *PhS* was the culmination of a widespread interest in philosophical history among Gotthold Lessing, Schiller, Goethe, and others of Hegel's contemporaries in which the themes of negative learning and historiographical reconstruction based on a biogenetic narrative play prominent roles; see Walter Kaufmann, *Hegel: A Reinterpretation* (Garden City, N.Y.: Anchor-Doubleday, 1959), 22–45.

22 See chapter 5 for examples from *P&R* of falsified mathematical history. The theme is an important one in Lakatos's historiographical theory.

23 One of Lakatos's best-known aphorisms, which he credits to Norwood Hanson, is "Philosophy of science without history of science is empty; history of science without philosophy of science is blind" (*HS*, 102). This historicized paraphrase of Kant's aphorism, "Concepts without intuitions are empty. Intuitions without concepts are blind," means also in *P&R* that the value of method is proportional to the history it informs, and that worthwhile history needs to be guided by a philosophically grounded conception of mathematical proof.

24 Georg Lukács, *Goethe and His Age,* trans. Robert Anchor (London: Merlin Press, 1968), 179. Lukács's essays here on *Wilhelm Meister,* Schiller, and *Faust* contain many observations relevant to the historiography of *PhS* and indirectly to *P&R*, such as Lukács's additional (pejorative for Schiller and Hegel, but epistemologically accurate for Lakatos) comment that Hegel's historiography "creates an iridescent twilight zone between history and logic" (129). Lakatos similarly characterizes heuristic as neither psychology nor logic; see *P&R*, 146.

25 Johann Wolfgang von Goethe, *Goethe's Faust,* trans. Walter Kaufmann (Garden City, N.Y.: Doubleday, 1961), 2. 4725.

26 As mathematical pedagogy, *P&R* extends George Pólya's revival of mathematical heuristic in Pólya's *How to Solve It, Mathematics and Plausible Reasoning,* and the advanced *Problems and Theorems in Analysis* written with George Szegö. The extension consists in Lakatos's heuristics for theorems to prove versus Pólya's heuristics for problems to solve. Where Pólya mainly investigates heuristics that apply a body of established techniques, such as proofs by induction, the evaluation of integrals or series, or the calculation of probability values, Lakatos's heuristics have as their aim to improve a naive conjecture into a genuine theorem, and possibly to introduce new concepts into mathematical theory.

27 On *The Critique of Judgment,* see Lukács, *The Young Hegel,* 343 ff.

28 Quoted in Lukács, *The Young Hegel,* 431. The criticism of such elitist class divisions is also central to Popper's account of the philosophical bases for totalitarianism in *The Open Society and Its Enemies.*

29 The irrationalist theme not only provides an important intellectual link to Lukács's Hegel, as already indicated, but to Lakatos's Hungary, too, as described in chapter 12.

2 | The Method of Proofs and Refutations

1 Lakatos also associates his notion of a fallible thought experiment with historical practices of Greek mathematics (*P&R*, 9 n. 1). On postulates and axioms in Greek mathematics as conjectural categories derived from dialectic, see Árpád Szabó,

The Beginnings of Greek Mathematics, trans. A. M. Ungar (Dordrecht, Netherlands: D. Reidel, 1978). Szabó's controversial thesis is that Greek mathematics became deductive and antiempirical through the outside influence of Eleatic philosophy in the fifth century B.C.E.

2 The proof of $V - E + F = 2$ that is central to *P&R* is Cauchy's, read in 1811 and published in 1813. For a summary of seven versions of $V - E + F = 2$ proved in chapter 1 of *P&R* (not including Poincaré's vector-algebraic proof of chapter 2), see Tuen Koetsier, *Lakatos' Philosophy of Mathematics: An Historical Approach* (Amsterdam: North-Holland, 1991), 28 ff. Proofs of the Euler(-Descartes) theorem existed before Cauchy, who gave the first nonmetrical proof—that is, one using only geometric properties. For a summary of the shortcomings of *P&R* as a comprehensive philosophy of mathematics, see Solomon Feferman, "The Logic of Mathematical Discovery vs. the Logical Structure of Mathematics," in *In the Light of Logic* (New York: Oxford University Press, 1999).

3 For the actual history, see *P&R*, 35 n. 2, 35 n. 4.

4 As an example of an important hidden lemma, one can try to dissect the disk-shaped faces of a cylinder with a diagonal, but if there are no vertices, how is the dissection performed? Is the possibility of dissection in that case "vacuously true," or is there an implicit existential clause that there is at least one diagonal cutting a planar face in two? In the end, the cylinder turns out not to satisfy the conditions for Poincaré's last modern proof provided in *P&R* and is excluded from the domain of Euler's theorem.

5 Quoted in Garrett Birkhoff, ed., *A Source Book in Classical Analysis* (Cambridge, Mass.: Harvard University Press, 1973), 61–62. For historical analysis on Seidel supporting Lakatos's account, see Ivor Grattan-Guinness, *The Development of Mathematical Analysis from Euler to Riemann* (Cambridge, Mass.: MIT Press, 1970), 112 ff., possibly influenced by Lakatos's ideas. For Lakatos's discussion of Seidel, see *P&R*, 136 ff. Seidel introduced the arbitrarily slow convergence of series of functions and not, as Lakatos says incorrectly, uniform convergence.

6 A sequence of functions $f_n(x)$ converges pointwise on (a,b) to $f(x)$ if for each $\epsilon > 0$ and $x \in (a,b)$ there is an $N_{\epsilon,x}$ such that $n > N_{\epsilon,x} \rightarrow |f_n(x) - f(x)| < \epsilon$. Uniform convergence interchanges the inner universal and existential quantifiers, and removes the dependence of N on x: for each $\epsilon > 0$ there is an N_ϵ such that for all $x \in (a,b)$, $n > N_\epsilon \rightarrow |f_n(x) - f(x)| < \epsilon$. An infinite series $\Sigma f_n(x)$ converges uniformly or pointwise according to whether the sequence of finite partial sums over M terms converges uniformly or pointwise. On pointwise and uniform convergence in nineteenth- and twentieth-century mathematics, see Grattan-Guinness, *Development,* 117 ff; on quantifier interchanges in other important theorems, see note 19 below.

7 The Dirichlet conditions are that the function to be expanded, taken as being defined on the interval $[-\pi, \pi]$, has a finite number of turning points, and is bounded and continuous, except perhaps for a finite number of finite discontinuities. The conditions preclude wild oscillations inside the interval.

8 On Abel, see Grattan-Guinness, *Development,* 84 ff. Abel's example is the Fourier sine series of $x/2$ for $0 \leq x \leq \pi/2$, $\sin x - 1/2 \sin 2x + 1/3 \sin 3x - \ldots$ as shown in Figure 4.

9 See Grattan-Guinness, *Development,* 110 ff.

10 "Starting from the certainty just achieved, that the theorem is not universally valid, and hence that its proof must rest on some extra hidden assumption, one then subjects the proof to a more detailed analysis. It is not very difficult to discover the hidden hypothesis. One can then infer backwards that this condition expressed by the hypothesis is not satisfied by series which represent discontinuous functions, since only thus can the agreement between the otherwise correct proof sequence, and what has been on the other hand established, be restored" (Seidel quoted in Birkhoff, *Source Book,* 62; see also *P&R*, 136).

11 See A. P. Youschkevitch, "The Concept of Function up to the Middle of the Nineteenth Century," *Archive for History of Exact Sciences* 16 (1976): 69.

12 See, for example, Carl Boyer, *The History of the Calculus and Its Conceptual Development* (New York: Dover, 1959), 282. After the logician Abraham Robinson used the compactness theorem of first-order logic to show that calculus could be rigorously based on Cauchy-like infinitesimals, some including Lakatos hoped that Cauchy's apparent lack of rigor could be explained via Robinson's nonstandard real number continuum. The attempt to rehabilitate Cauchy did not pan out; see introduction, note 23.

13 In terms of what came before the nineteenth century, the mathematical world was quite different again. Mathematics was not clearly delimited from physics and other sciences around 1800, physical arguments were often put forth in mathematical proofs, and the process whereby mathematics was purified of physical and metaphysical doctrines was dependent on specific mathematical discoveries making them unnecessary. Not until 1854 would Riemann explicitly recognize the need to extend mathematical analysis beyond functions "found in nature." Such changes go beyond those in general standards of criticism, like Lakatos's, that are not mathematically specific. See J. R. Ravetz, "Vibrating Strings and Arbitrary Functions," in *The Logic of Personal Knowledge: Essays Presented to Michael Polanyi on His Seventieth Birthday,* ed. M. Grene (Glencoe, Ill.: Free Press, 1961).

14 Salomon Bochner, *The Role of Mathematics in the Rise of Science* (Princeton, N.J.: Princeton University Press, 1966), 330. Lagrange was known for his austerity but the Greeks already valued multiple proofs highly; see, for instance, the several proofs of the "Delian problem" of doubling the cube in Thomas Heath, *A History of Greek Mathematics* (New York: Dover, 1981), 1: 244–70.

15 See Grattan-Guinness, *Development,* 50. Grattan-Guinness cannot date a "change in intent," capturing the definition's geometric implications, earlier than 1844.

16 Hegel's quote is Walter Kaufmann's translation of "Die Meinung erfahrt, dass es anders gemeint ist, als sei meinte." Similarly, "What is familiar is not known just because it is familiar," is Kaufmann's version of "Das Bekannte überhaupt ist darum, weil es bekannt ist, nicht erkannt" (*Pf*, 48).

17 Hegel provides a lengthy discussion on mathematics in his preface (*Pf*, 62–69). It is tempting to see Lakatos as refuting Hegel's distinction between temporal reason and static mathematical understanding, but that would be incorrect as explained in the text. See also chapter 1, note 19 above.

18 Hegel, from 1787, quoted in H. S. Harris, *Hegel's Development: Towards the*

Sunlight, 1770–1801 (New York: Oxford University Press, 1972), 19. On language in Hegel, see Michael Forster, *Hegel's Idea of a Phenomenology of Spirit* (Chicago: University of Chicago Press, 1998), 83, 205.

19 The logical error in Cauchy's false continuity theorem involves the interchange of a universal and an existential quantifier in the definitions of pointwise and uniform convergence. Logically the error is deep, for a similar interchange occurs, under specified conditions, in the Heine-Borel theorem, the related compactness theorem for first-order logic, and the weak and strong laws of large numbers in probability theory. But the structural affinities between these theorems are made possible by knowing how to interpret the informal proofs so that formal similarities are revealed: the theoretical notion of a quantifier-interchange seems natural today, but not to all mathematicians of the nineteenth century, even if they recognized specific instances. See also this chapter's note 6.

20 Ivor Grattan-Guinness, "The Emergence of Mathematical Analysis and Its Foundational Progress, 1780–1880," in *From the Calculus to Set Theory: 1630–1930* (London: Duckworth, 1980), 139. For a Lakatosian discussion of "different proofs prove different theorems" by Grattan-Guinness, see also ibid., 146 ff.

21 Henri Lebesgue, still interested in the problem in 1905, used $\lim_{m\to\infty} [\lim_{n\to\infty} (\cos m!\,\pi x)^{2n}]$ to show that "it is not evident that there are functions which cannot be analytically represented" (quoted in A. F. Monna, "The Concept of Functions in the Nineteenth and Twentieth Centuries, in Particular with Regard to the Discussions between Baire, Borel, and Lebesgue," *Archive for History of Exact Sciences* 9 [1972]: 71). Birkhoff credits Dirichlet with this expression in *Source Book,* 70 n. 8. Hankel, in 1870, provided an expression for $\boldsymbol{Q}(x)$ incorporating a Fourier sine series, contrary to Dirichlet's belief that the function could not be so expressed; see Grattan-Guinness, "Emergence of Mathematical Analysis," 139.

22 On the history of Hankel's work, see Thomas Hawkins, *Lebesgue's Theory of Integration: Its Origins and Development* (Madison: University of Wisconsin Press, 1970), 28–59.

23 Cantor, in 1880 looking back on 1872, quoted in Joseph Dauben, "The Trigonometric Background to Georg Cantor's Theory of Sets," *Archive for History of Exact Sciences* 7 (1971): 214. On Cantor's creation of novel mathematical concepts, Dauben concludes: "It should be stressed that the concept of limit-point and the associated idea of the first derived set P' of limit-points of a set P are the fundamental, basic elements of Cantor's development of the set theory" (209 n. 65). On new objects, explains Dauben, the mark of Cantor's work was "creating new forms and concepts when the limitations of existing approaches here became evident . . . [and] forging new concepts and tools which could be used to obtain radically new kinds of solutions to known problems" (216).

24 More precisely, if Zermelo-Fraenkel set theory is consistent, then so is an extension in which all sets of real numbers are Lebesgue measurable. In Zermelo-Fraenkel set theory plus the axiom of choice, some sets of real numbers can be proved not to be Lebesgue measurable.

25 Hadamard's comment occurred in a famous 1904 exchange of letters between himself, Baire, Borel, and Lebesgue over the axiom of choice, translated in Greg-

ory Moore, *Zermelo's Axiom of Choice: Its Origins, Development, and Influence* (New York: Springer, 1982), 318; in the fourth of the five letters.

26 See Stephen Simpson, *Subsystems of Second Order Arithmetic* (New York: Springer, 1999); and Solomon Feferman, *In the Light of Logic* (New York: Oxford University Press, 1999).

3 | Mathematical Skepticism

1 After working in secret for about seven years, but having had his proof confidentially reviewed by another world-class mathematician, Andrew Wiles publicized his first proof of Fermat's last theorem in June 1993. The proof contained an irreparable error. While fatal to the first approach, however, the error led after about eighteen months to Wiles's understanding of how he could successfully use an original strategy discarded years before. The case is a spectacular example of a falsified proof and lemma-incorporation. Still, nothing guarantees the proof's correctness today.

2 On the development of first-order logic as a dominant logical theory, see chapter 4, note 5 below.

3 See Thomas Hawkins, *Lebesgue's Theory of Integration: Its Origins and Development* (Madison: University of Wisconsin Press, 1970); and Joseph Dauben, *Georg Cantor: His Mathematics and Philosophy of the Infinite* (Princeton, N.J.: Princeton University Press, 1979), chapters 1–2.

4 Árpád Szabó, *The Beginnings of Greek Mathematics,* trans. A. M. Ungar (Dordrecht, Netherlands: D. Reidel, 1978), 177.

5 The role of skepticism as a piece of Hegel's critical method has long been noted, but only recently has its deeper role been explored by Michael Forster in *Hegel and Skepticism* (Cambridge, Mass.: Harvard University Press, 1989). My suggestion that the phenomenological method is effectively nothing but transmogrified tropes is outlined in "A Mathematical Bildungsroman," *History and Theory* 28 (1989): 36 n. 33, and differs from Forster's approach. In any case, it is now clear that the widely accepted notion that the role of skepticism in modern philosophy ends with Kant is in need of revision. Skepticism continues past Kant directly through Hegel, and then in mediated form, to Marx and critical theory; Marx criticized much in Hegel, but nonetheless found in *PhS* the germ of a powerful critical and historical technique (see, for example, Karl Marx, "Critique of Hegel's Dialectic and General Philosophy," in *Karl Marx: Early Writings,* trans. and ed. Tom Bottomore (New York: McGraw-Hill, 1963). For an example of standard historiography, see Myles Burnyeat's introduction to *The Skeptical Tradition* (Berkeley: University of California Press, 1983): "In philosophical writing after Kant 'skepticism' and 'the skeptic' increasingly become schematic, ahistorical notions. . . . [Barry] Stroud's essay [in the volume] explains why Kant brought the skeptical tradition to an end. Not that there is nothing left for contemporary philosophers to say about various kinds of skepticism. But the skepticisms they are talking about are a free creation of the modern philosophical imagination. They no longer descend from the ancient lineage of Pyhrro and the Academy" (3).

6 Pyrrhonism is named for the legendary Pyrrho of Elis, who lived circa 360–275

B.C.E. Pyrrho did not create any formal doctrine but was thought to continuously suspend judgment on moral and ethical beliefs as a consistent living practice. Versions of theoretical skepticism are associated with Arcesilaus, Carneades, and Aenesidemus from the fourth through the second centuries B.C.E., with the latter and his followers known for their formulation of Pyrrhonism as a series of tropes or dialectical techniques for bringing arguments to various conclusions. Arcesilaus and Carneades are sometimes characterized as negative dogmatists for reaching stronger conclusions regarding what *cannot* be known, in contrast to the weaker suspension of judgment of the later Pyrrhonists. Sextus Empiricus's *Outlines of Pyrrhonism* (trans. R. C. Bury, 4 vols. [Cambridge, U.K.: Loeb Classical Library, 1939]) provides the only full treatment of Pyrrhonist technique and the tropes, and summarizes some five hundred years of skeptical method along with philosophies taken as the object of skeptical attack.

7 On ancient versus modern skepticism for Hegel, see Forster, *Hegel and Skepticism,* chapters 1–2. Hegel's account is shown by Forster to be largely correct and not an eccentric interpretation. On various roles for skepticism in German philosophy between Kant and Hegel, see George di Giovanni and H. S. Harris, *Between Kant and Hegel: Texts in the Development of Post-Kantian Idealism* (Albany: State University of New York Press, 1985). A typical characterization of Hegel on Hume's skepticism is that it "should be clearly marked off from Greek skepticism. Hume assumes the truth of the empirical element, feeling and sensation, and proceeds to challenge universal principles and laws, because they have no warranty from sense-perception. So far was ancient skepticism from making feeling and sensation the canon of truth, that it turned against the deliverances of sense first of all" (*Hegel's Logic,* pt. 1 of *Hegel's Encyclopedia,* trans. William Wallace [Oxford: Oxford University Press, 1975], 64).

8 Sextus, *Outlines of Pyrrhonism,* 1: 7–9.

9 Ibid., 1: 163–65.

10 Ibid., 2: 239.

11 Ibid., 2: 87. This is Charlotte Stough's correction to Bury's translation of Sextus as noted in her *Greek Skepticism* (Berkeley: University of California Press, 1969), 58. See also G. W. F. Hegel, *Lectures on the History of Philosophy,* trans. E. S. Haldane and F. H. Simson (New York: Humanities Press, 1974): "But there is a misunderstanding or a formal understanding in considering that all philosophy that is not Scepticism is Dogmatism" (2: 363). As suggested by Stough's example, some distinctions needed to state Pyrrhonist claims without overstatement can be cast in terms of scope definitions of negations and quantifiers in the implicit logical form of skeptical statements.

12 In "The Skeptic's Two Kinds of Assent and the Question of the Possibility of Knowledge," Michael Fried discusses the skeptical practice of immanent critique, or developing critical arguments using logic and concepts derived from the opponent's position (in *Philosophy in History: Essays on the Historiography of Philosophy,* ed. Richard Rorty et al. [New York: Cambridge University Press, 1984], 258–59). This technique as mediated through Hegel is central to the Young Hegelians, Marx, and then critical theory.

13 Sextus, *Outlines of Pyrrhonism,* 2: 83. See also Pierre Couissin, "The Stoicism of

the New Academy," in *The Skeptical Tradition,* ed. Myles Burnyeat (Berkeley: University of California Press, 1983), 41; and Sextus, *Outlines of Pyrrhonism,* on Clitomachus and his followers: "For by plunging into alien subject matter and framing their arguments on the basis of assent to dogmatic assumptions not their own they have unduly prolonged their counter-statement" (3: 3).

14 See David Sedly, "The Motivation of Greek Skepticism," in *The Skeptical Tradition,* ed. Myles Burnyeat (Berkeley: University of California Press, 1983), 17.

15 See Myles Burnyeat, "The Sceptic in His Place and Time," in *Philosophy in History: Essays on the Historiography of Philosophy,* ed. Richard Rorty (New York: Cambridge University Press, 1984), 243.

16 See Myles Burnyeat, "Can the Skeptic Live His Skepticism?" in *The Skeptical Tradition,* ed. Myles Burnyeat (Berkeley: University of California Press, 1983), 142–43 n. 8.

17 Ibid. For instance: "As for *to phainomenon,* what appears may, so far as I can see, be anything whatever. . . . [T]he skeptic contrast between appearance and real existence is a purely formal one, entirely independent of subject matter" (127–28).

18 Richard Popkin, *The History of Scepticism: From Erasmus to Spinoza,* 3d ed. (Berkeley: University of California Press, 1979), 15.

19 Ibid., 126.

20 David Hume, *Dialogues Concerning Natural Religion* (Indianapolis, Ind.: Bobbs-Merrill, 1947), 132.

21 David Hume, *A Treatise of Human Nature* (Buffalo, N.Y.: Prometheus, 1992), 267.

22 Feyerabend also noted that his arguments were sometimes predicated only on the rationalist discourse assumed by contemporary philosophers of science; on his skepticism, see Paul Feyerabend, *Against Method* (London: New Left Books, 1975), 221 ff.

23 Hume, *A Treatise of Human Nature,* 186.

24 Quoted in Richard Popkin, "David Hume and the Pyrrhonian Controversy," in *The High Road to Pyrrhonism,* ed. R. A. Watson and J. E. Force (San Diego, Calif.: Austin Hill Press, 1980), 133.

25 Ibid.

26 Popkin's discussion of Humean-mitigated skepticism as "schizophrenia" is in his "David Hume: His Pyrrhonism and His Critique of Pyrrhonism," in *The High Road to Pyrrhonism,* ed. R. A. Watson and J. E. Force (San Diego, Calif.: Austin Hill Press, 1980). For Hume on mitigated skepticism, see his *An Enquiry concerning Human Understanding,* ed. L. A. Selby-Bigge and P. H. Nidditch, 3d ed. (Oxford: Oxford University Press, 1975), 161 ff. In my account, I am not trying to characterize Hume per se but rather use him as a voice for the ideas presented on unmitigated theoretical skepticism. Hume's own mitigated skepticism included a kind of empirical naturalism, and is described in Robert Fogelin, "The Tendency of Hume's Skepticism," in *The Skeptical Tradition,* ed. Myles Burnyeat (Berkeley: University of California Press, 1983).

27 Hume, *Dialogues Concerning Natural Religion,* 219 n.

28 Sextus, *Outlines of Pyrrhonism,* 1: 135.

29 See Gisela Striker, "The Ten Tropes of Aenesidemus," in *The Skeptical Tradition,* ed. Myles Burnyeat (Berkeley: University of California Press, 1983).

30 Separation of the "we" as readers of *PhS* from the consciousness "under view," and then merging them, is one of the book's primary narrative methods. See Kenley Dove, "Hegel's Phenomenological Method," in *New Studies in Hegel's Philosophy,* ed. Warren Steinkraus (New York: Holt, Rinehart and Winston, 1971). In *P&R*, the separation is handled via the footnote apparatus and dialogue characters.

31 On the two interpretations of ou mallon, see Striker, "The Ten Tropes of Aenesidemus," 112.

32 On self-reference in Pyrrhonism, see Myles Burnyeat, "Protagoras and Self-Refutation in Later Greek Philosophy," *Philosophical Review* 85 (1976): 53. The self-referential pointing back to earlier stages of argument by the skeptic is also central to Gödel's second incompleteness theorem as discussed in chapter 4.

33 For the remainder of this chapter, the account is my own, not Lakatos's, as is the material on Gödel in chapter 4.

34 See note 8 above and the text for Sextus on skeptical ability.

35 There have been many philosophical debates on the "truth" of the axioms of constructability, determinacy, projective determinacy, the generalized continuum hypothesis, and so on, with no choice appearing compelling on mathematical grounds: there are many options with advantages and disadvantages, but no prospect of a winner. That could one day change, of course, although at present these various axiom systems appear like so many exotic geometries or algebras. For a discussion, see Penelope Maddy, "Believing the Axioms, I, II," *Journal of Symbolic Logic* 53 (1988): 481–511, 736–64; and chapter 4, note 9, and text below.

36 Quoted in Pierre Duhem, *The Aim and Structure of Physical Theory,* trans. Philip Weiner (Princeton, N.J.: Princeton University Press, 1954), 27. Likewise, as Pierre Bayle observed: "I know too much to be a Pyrrhonist, and I know too little to be a dogmatist" (quoted in Popkin, *History of Scepticism,* 213). For a similar expression of "pragmatic skepticism," citing John Dewey as a precedent, see the conclusion in *Scientific Practice: Theories and Stories of Doing Physics,* ed. Jed Buchwald (Chicago: University of Chicago Press, 1995), 345–46.

4 | Between Formal and Informal

1 The scare quotes are a reminder that a basic message of *P&R* is that no part of mathematics is immune to reinterpretation (thus there are no perfectly known terms) or refutation (no inferences are infallible). The chapter and appendix organization of *P&R* is the product of Lakatos's editors. Continuity of the texts was apparently indicated to them by Lakatos. For criticism of Worrall and Zahar's editorial comments on logic and fallibility, see Philip Davis and Reuben Hersh, *The Mathematical Experience* (Boston: Houghton Mifflin, 1981), 353–56.

2 Recall the proof is Poincaré's; see *P&R*, 52, where he speaks this line via Alpha.

3 Epsilon does not really carry his project to completion; he trades on ambiguities of his presentation to allow some informal meaning within his supposedly purely formal calculus, since no informal proof is completely formalized: steps are skipped and some consequences turn out to be explained heuristically, or perhaps by example. Georg Kreisel's comment on formal rigor is relevant: "The widespread idea of connecting formalization and reliability is equally incoherent. When we really want to make sure that a mathematical result is correct we do not

formalize its proof, i.e., compare the steps with given formal rules, but we try to make it intelligible. We rely on mathematical, not logical foundations!" ("A Survey of Proof Theory," *Journal of Symbolic Logic* 33 [1968]: 361). On this point, see also Paul Feyerabend, *Against Method* (London: New Left Books, 1975), 183 n. 7.

4 According to Plato in the *Republic,* it was difficult for many to be clear about the new, antiempirical, intellectualist foundation for mathematical geometry. Locutions such as "squaring" or "cubing" were misleading for Plato because geometry was not supposed to have any relation to real or perceived figures. Like arithmetic, which was prized by the Greeks as the purer science, geometry should be a science of abstractions based on logical rules. Plato complained that "geometry is in direct contradiction with the language employed in it by its adepts . . . they speak as if they were doing something [practical]. . . . [T]heir talk is of squaring and applying and adding and the like, whereas in fact the real object of the study is pure knowledge." (*Republic,* in *The Collected Dialogues,* ed. Edith Hamilton and Huntington Cairns [Princeton, N.J.: Princeton University Press, 1961], 527a, 759). Geometric mathematics in the time of Plato was an impure science of space. See Árpád Szabó, *The Beginnings of Greek Mathematics,* trans. A. M. Ungar (Dordrecht, Netherlands: D. Reidel, 1978), 307 ff.

5 On Thoralf Skolem's role in making first-order logic a dominant theory for parts of modern logic, see Gregory Moore, "The Emergence of First-Order Logic," in *History and Philosophy of Modern Mathematics,* ed. William Aspray and Philip Kitcher (Minneapolis: University of Minnesota Press, 1988). On the same development and its completion with Gödel, see Warren Goldfarb, "Logic in the Twenties: The Nature of the Quantifier," *Journal of Symbolic Logic* 44 (1979): 351–68.

6 See *P&R*, 2 n. 2; and *MSE*, 37.

7 Gödel's argument cannot be carried out without axioms for both addition and multiplication, essentially because Gödel numbering requires exponentiation defined using both. The theories of the natural numbers with only the operation of multiplication or addition are both decidable: the sets of first-order sentences true in the structures $<\boldsymbol{N}, \times>$ and $<\boldsymbol{N}, +>$ are both recursive. One part of Gödel's first theorem also uses ω-consistency rather than simple consistency. The term "system" here means a recursive set of axioms used to generate its theorems using logical rules.

8 See Jeff Paris and Leo Harrington, "A Mathematical Incompleteness in Peano Arithmetic," in *Handbook of Mathematical Logic,* ed. Jon Barwise (Amsterdam: North-Holland, 1977).

9 It was Gödel, in 1947, who originally suggested that the new axioms investigated by set theorists could be justified by their lower-order mathematical consequences in "What Is Cantor's Continuum Problem?" in *The Philosophy of Mathematics,* ed. Paul Benacerraf and Hilary Putnam (Englewood Cliffs, N.J.: Prentice-Hall, 1964).

10 *The Theaetetus of Plato,* trans. M. J. Levett and Myles Burnyeat (Indianapolis, Ind.: Hackett, 1990), 171a (298).

11 In setting out to prove the second theorem, Cohen, in describing the full and final arithmetization, says that "this is perhaps the most subtle point in the entire

subject and the reader is urged to stop and review carefully what we have done up to here" (Paul Cohen, *Set Theory and the Continuum Hypothesis* [Reading, Pa.: W. A. Benjamin, 1966], 42).

12 John Dawson argues that it is a bit of a myth that Gödel's results were accepted without ado. Rather, "there were doubters and critics, as well as defenders and rival claimants to priority" ("The Reception of Gödel's Incompleteness Theorems," in *Proceedings of the 1984 Biennial Meeting of the Philosophy of Science Association,* ed. Peter Asquith and Philip Kitcher [East Lansing, Mich.: Philosophy of Science Association, 1985], 254).

13 One might say: assuming the consistency of the arithmetic system. But if the system is inconsistent, all statements are provable and the two formulations are again coextensive.

14 See the references to Andrzej Mostowski and Alan Turing in Solomon Feferman, "Arithmetization of Metamathematics in a General Setting," *Fundamenta Mathematicae* 49 (1960): 35–92.

15 For an example, see theorem 3 in G. T. Kneebone, *Mathematical Logic and the Foundations of Mathematics* (London: Van Nostrand, 1963), 240. On falsification of mathematical history in *P&R*, see chapter 5, note 8, and text below.

16 Georg Kreisel discusses the problem of translation vis-à-vis the incompleteness theorems in "Mathematical Significance of Consistency Proofs," *Journal of Symbolic Logic* 23 (1958): 177, and "A Survey of Proof Theory," 323. Kreisel's note on Henkin's problem is "On a Problem of Henkin's," *Indagationes Mathematicae* 15 (1953): 405–6. Kreisel was actually unsympathetic to the study of noncanonical consistency statements, believing that a canonical choice was required to generalize Gödel's second incompleteness theorem. See Solomon Feferman, "My Route to Arithmetization," *Theoria* 63 (1997), 178. Feferman also discusses unexpected applications of nonstandard consistency formulations in proof theory.

17 See Feferman, "Arithmetization of Metamathematics." Feferman also provides a list of results that he classifies as intensional or extensional. The result, writes Feferman, is "*extensional* if essentially only numerically correct definitions are needed, or *intensional* if the definitions must more fully express the notions involved, so that various of the general properties of these notions can be formally derived" (35). Feferman's approach is to use the standard consistency sentence, but modify the axioms from which they are proved, instead of vice versa, as described in the text.

18 Ibid., 67–68.

19 Ibid., 79.

20 The provability conditions, with "$\Box$" intended as "it is provable that," are: $\Box(\boldsymbol{p} \rightarrow \boldsymbol{q}) \rightarrow \Box\boldsymbol{p} \rightarrow \Box\boldsymbol{q}$; $\Box\boldsymbol{p} \rightarrow \Box\Box\boldsymbol{p}$; and $\Box(\Box\boldsymbol{p} \rightarrow \boldsymbol{p}) \rightarrow \Box\boldsymbol{p}$. The last condition is the abstraction of Löb's theorem: if it is provable that *a proof of* $\boldsymbol{p}$ implies $\boldsymbol{p}$, then $\boldsymbol{p}$ is indeed provable. The connection between provability and modal logics was pointed out for the first time by Gödel in "An Interpretation of the Intuitionistic Sentential Logic" (1933), in *Kurt Gödel: Collected Works, Volume I,* ed. Solomon Feferman et al. (New York: Oxford University Press, 1986), In Stephen Kleene's notes to Gödel's 1931 undecidability paper in *Collected Works, I,* he writes that a detailed proof of the second theorem was planned for the sequel, but never writ-

ten, "in part because Gödel felt that the results of Part I had won prompt acceptance. [But see note 12 above.] Certainly the idea of the argument for [the second theorem] was very convincing; but it turned out that the execution of the details required somewhat more work and care than had been anticipated. . . . A demonstration that [the second theorem] is provable was given in Hilbert and Bernays 1939" (137). Gödel's definition of consistency is provided in "On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems I," in *Collected Works, I,* 193 n. 63.

21 George Boolos, *The Unprovability of Consistency* (New York: Cambridge University Press, 1979), 137.

22 Boolos himself remarks that difficult theoretical problems with Rosser sentences may turn on the "(awful? attractive?) possibility that . . . the answers . . . may depend on the way sentences of arithmetic are assigned gödel numbers" (ibid., 139).

5 | Reason Inverted

1 Quoted in Thomas Heath, *A History of Greek Mathematics* (New York: Dover, 1981), 2: 21.

2 See Jaakko Hintikka and Unto Remes, *The Method of Analysis: Its Geometrical Origin and Its General Significance* (Boston: D. Reidel, 1974), 105; on Galileo and Descartes, see Heath, *A History of Greek Mathematics,* ibid.

3 "*Analysis,* then, takes that which is sought as if it were admitted and passes from it through its successive consequences to something which is admitted as the result of synthesis: for in analysis we assume that which is sought as if it were already done, and we inquire what it is from which this results, and again what is the antecedent cause of the latter, and so on, until by so retracing our steps we come upon something already known or belonging to the class of first principles, and such a method we call analysis as being solution backwards. But in *synthesis,* reversing the process, we take as already done that which was last arrived at in the analysis and, by arranging in their natural order as consequences what before were antecedents, and successively connecting them one with another, we arrive finally at the construction of what was sought; and this we call synthesis" (Heath, *A History of Greek Mathematics,* 2:400).

4 Aristotle, *Nicomachean Ethics,* book 3, 1112b, 15–25, in *The Complete Works of Aristotle,* ed. Jonathan Barnes (Princeton, N.J.: Princeton University Press, 1984), 2: 1756.

5 Quoted in *P&R*, 107.

6 See Kurt Gödel, "What Is Cantor's Continuum Problem?" in *The Philosophy of Mathematics,* ed. Paul Benacerraf and Hilary Putnam (Englewood Cliffs, N.J.: Prentice-Hall, 1964): "But, despite their remoteness from sense experience, we do have something like a perception also of the objects of set theory, as is seen from the fact that the axioms force themselves upon us as being true. I don't see any reason why we should have less confidence in this kind of perception, i.e., in mathematical intuition, than in sense perception" (271–72). In fairness to Gödel, it was mostly philosophers who took up his Platonism, while mathematicians were influenced by his approach to judging various set theories by their consequences, and the latter is the principal topic of Gödel's Cantor paper.

7 Quoted in *MSE*, 35, from Gödel's original publication of his proof of the consistency of the continuum hypothesis and axiom of choice. See also there the criticism of Gödel's "logical optics" by Hermann Weyl.
8 On Abel and uniform convergence, see *P&R*, 134–35. Nicolas Bourbaki is the collective pseudonym for a distinguished set of French mathematicians who published many books under that name. See also the errors in accounts of Gödel's theorems in chapter 4, note 15, above. On falsification of economic history in Marx, see chapter 11 below.
9 Karl Marx, *Grundrisse: Foundations of the Critique of Political Economy,* trans. Martin Nicolaus (New York: Vintage, 1973), 100–108.
10 Ibid., 58, 60.
11 Ibid., 60. Other discussions of the order of discovery and presentation occur in Karl Marx, *Capital,* ed. Friedrich Engels, trans. Samuel Moore and Edward Aveling, 3d ed. (New York: International Publishers, 1967), 1: 19ff. (in the afterword to the second German edition), and in "The Fetishism of Commodities and the Secret Thereof"—for example, "Man's reflections on the forms of social life, and consequently, also, his scientific analysis of those forms, take a course directly opposite to that of their actual historical development" (1: 75ff.).
12 See introduction, note 12, above.
13 Karl Popper, *Objective Knowledge: An Evolutionary Approach* (Oxford: Oxford University Press, 1972), chapter 3. Popper compares his approach to Hegel (125–26), where he is misdescribed by Popper as a Platonist or neo-Platonist.
14 "The forms of thought are, in the first instance, displayed and stored in human *language*" (G. W. F. Hegel, *Hegel's Science of Logic,* trans. A. V. Miller [New York: Humanities Press, 1976], 31). On Hegel's development of new meanings for common words in line with his conception of Bildung, see Michael Forster, *Hegel's Idea of a Phenomenology of Spirit* (Chicago: University of Chicago Press, 1998), 104 n. 212; on a similar process for Johann Gottfried von Herder, see ibid., 110 n. 228. The process applies well to Lakatos and terms like his "lemma-incorporation," "research programme," and so on.
15 Discussion of linguistic objectivity could also include Marx's theories of objectification and alienation, but would require considerable caveats because of Marx's almost completely nonlinguistic approach.
16 See Paul Feyerabend, *Problems of Empiricism: Philosophical Papers, Volume 2* (New York: Cambridge University Press, 1981), chapter 9. For a related account of literacy and the alphabet in Greek mathematics, see my "Dialectic and Diagonalization," *Inquiry* 34 (1991): 3–25.

6 | Kuhn, Popper, Feyerabend, Lakatos

1 See Michael Forster, *Hegel's Idea of a Phenomenology of Spirit* (Chicago: University of Chicago Press, 1998), 366–409, for a critique of Donald Davidson on conceptual schemes. For an uncritical acceptance of Davidson's ideas as an answer to problems of conceptual change, see John Preston, *Feyerabend: Philosophy, Science, and Society* (Cambridge, U.K.: Polity Press, 1997), 189.
2 Paul Feyerabend, *Against Method* (London: New Left Books, 1975), 17. Feyerabend is quoting Lenin from "Left-Wing Communism—An Infantile Disorder,"

with changes such as "best methodologist" instead of "best parties." That opening and its clever quotation is fitting given that *Against Method* is written as "a long and rather personal letter" to Lakatos (dedication). See also introduction, note 5, above.

3 On misinterpretations of *Against Method,* see Paul Feyerabend, *Killing Time: The Autobiography of Paul Feyerabend* (Chicago: University of Chicago Press, 1995), chapter 12.

4 Paul Feyerabend, *Problems of Empiricism: Philosophical Papers, Volume 2* (New York: Cambridge University Press, 1981), 12 ff.

5 See Paul Feyerabend, "Consolations for the Specialist," in *C&GK*, reprinted also as Feyerabend, *Problems of Empiricism,* chapter 8.

6 See chapter 7, note 32, and text below on propaganda roles in Kuhn's *Copernican Revolution.*

7 See Feyerabend, *Problems of Empiricism,* 25–33; and Paul Feyerabend, *Science in a Free Society* (London: New Left Books, 1978).

8 The journalist Harold Stearns said long ago that "the technique of liberal failure" was "the unwillingness of the liberal to continue with analysis once the process of analysis had become uncomfortable" (quoted in James Wechsler, *The Age of Suspicion* [New York: Random House, 1953], 39). On the other hand, Feyerabend was widely perceived as reckless for daring to criticize institutionalized science, and Lakatos was similarly seen by some as a demagogue (see chapter 9, note 12, and text).

9 See the essays by Popper and John Watkins in *C&GK*.

10 Thomas Kuhn, "Logic of Discovery or Psychology of Research?" in *C&GK*, 21. Lakatos comments on Kuhn's polar choices in *FM*, 10, 90–91. For a discussion of false implicit dichotomies in Kuhn and others' philosophies of science, see Peter Galison, "Context and Constraints," in *Scientific Practice: Theories and Stories of Doing Physics,* ed. Jed Buchwald (Chicago: University of Chicago Press, 1995), 29 ff.

11 Paul Feyerabend, "Imre Lakatos," *British Journal for the Philosophy of Science* 26 (1975): 6.

7 | A Historiographical Toolkit

1 Lakatos made one significant change in his ideas as presented in *FM* concerning the characterization of "novel facts" in a paper on Copernicus written with Elie Zahar and published in *MSRP*, chapter 4.

2 The "cues" are: *FM*, 30 (second par.), 31 n. 2, 32 (last par.), 52–53, 71 n. 2, 89 n. 5, 91, 92 n. 3.

3 "The second more profound form of negation consists in the unification of diverse philosophies into one whole in such a way that no one of them remains independent, but all appear as parts of the one. Their principles are united by being reduced to elements of the one idea. . . . The history of philosophy is the history of free, concrete thought—which is to say, of reason. Free concrete thought is concerned only with itself. . . . [T]he history of philosophy is identical with the system of philosophy . . . [and] is itself a science. . . . [P]hilosophy develops through

its history, and vice versa. Philosophy and history of philosophy mirror each other" (G. W. F. Hegel, introduction to *Lectures on the History of Philosophy,* in *Hegel's Idea of Philosophy,* ed. and trans. Quentin Lauer [New York: Fordham University Press, 1974], 88–89, 71, 83–84). See also G. W. F. Hegel, *Hegel's Logic,* pt. 1 of *Hegel's Encyclopedia,* trans. William Wallace (Oxford: Oxford University Press, 1975), sections 14, 19; and *H&CC*, 140.

4 See also *MSRP*, 153.

5 The term "game" is appropriate because the entire account is framed as criticism and has almost no traditional epistemological content. See *HS*, 122; *MSRP*, chapter 3, on Popper and problems of "verisimilitude"; and *LSCD* on methodology as conventional rules of "the game of empirical science" (53).

6 Little time is spent on theory ladenness in Lakatos, and one needs to look to Kuhn, Feyerabend, Norwood Hanson's *Patterns of Discovery* (New York: Cambridge University Press, 1969), or elsewhere for details.

7 As Lakatos says coyly, "If we want to falsify our methodological falsificationism, we have to do it before having a theory of how to do it" (*FM*, 30; see also *HS*, 130 n. 4). "Later," in the historiographical theory, the history used to falsify Popper_1 will also be theory-laden, just as facts are theory-laden in Popper_1.

8 The insertions are Lakatos's. This is about the only serious methodological comment on Marxism in Lakatos's work, and one that will be pursued in chapter 11 below.

9 Popper partially introduced the idea of progress without refutation, but Lakatos makes substantially more of it than did Popper.

10 For varied applications of the methodology of scientific research programmes, see Luce Giard, introduction to *Histoire et Methodologie des Sciences: Programmes de recherche et reconstruction rationnelle* (Paris: Presses Universitaires de France, 1994), xxxiii, n. 1. For applications in mathematics, see Teun Koetsier, *Lakatos' Philosophy of Mathematics: An Historical Approach* (Amsterdam: North-Holland, 1991), chapter 2; and Michael Hallet, "Towards a Theory of Mathematical Research Programmes," *British Journal for the Philosophy of Science* 30 (1979): 1–25, 135–59. In economics, see the books edited by Roger Backhouse and Spiro Latsis in the bibliography.

11 See Alan Musgrave, "Method or Madness: Can the Methodology of Scientific Research Programmes Be Rescued from Epistemological Anarchism?" in *Essays in Memory of Imre Lakatos,* ed. Robert Cohen et al. (Boston: D. Reidel, 1976). Musgrave also concludes that attributions like a "hard core" should only be made retrospectively in the spirit of the historiographical toolkit approach provided here.

12 Lakatos and Lakatos-Zahar provide a variety of characterizations of ad hoc theories; see Elie Zahar, "Why Did Einstein's Programme Supersede Lorentz's?" in *Method and Appraisal in the Physical Sciences: The Critical Background to Science, 1850–1900,* ed. Colin Howson (New York: Cambridge University Press, 1976), 216.

13 For excellent discussions of contradictions and hindsight, see, respectively, *SSR*, 79, 103. In "Reflections on My Critics," Kuhn provided a somewhat different account of hindsight after he saw Lakatos's treatment; see *C&GK*, 262.

14 See Colin Howson, ed., *Method and Appraisal in the Physical Sciences: The Critical Background in Science, 1850–1900* (New York: Cambridge University Press, 1976). The book includes a critical essay by Paul Feyerabend called "On the Critique of Scientific Reason," reprinted as *Problems of Empiricism: Philosophical Papers, Volume 2* (New York: Cambridge University Press, 1981), chapter 10. All the studies save for Martin Frické's on Avogadro's law are summarized here. Much detail is omitted to convey the major methodological points on the use of research programme ideas.

15 On research programmes as capturing part of the motivation behind Kuhn's normal science, see *FM*, 90 n. 1.

16 Except for the Lakatos-Zahar study of Copernicus (*MSRP*, chapter 4), the following studies are from Howson, *Method and Appraisal in the Physical Sciences:* John Worrall, "Thomas Young and the 'Refutation' of Newtonian Optics: A Case-Study in the Interaction of Philosophy of Science and History of Science"; Alan Musgrave, "Why Did Oxygen Supplant Phlogiston? Research Programmes in the Chemical Revolution"; Peter Clark, "Atomism versus Thermodynamics"; and Elie Zahar, "Why Did Einstein's Programme Supersede Lorentz's?"

17 See, for example, Robert Purrington, *Physics in the Nineteenth Century* (New Brunswick, N.J.: Rutgers University Press, 1997): "After a century of Newton's dominance, the [corpuscular theory] was so firmly entrenched that, when Thomas Young advanced his wave theory in 1801, the hostile reaction drove him out of optics" (39). Another explanation is that it was largely Young's explicit assumption of an ether for his waves to travel in that raised significant objections.

18 For Fresnel's version of Young's achievements, see Emilio Segrè, *From Falling Bodies to Radio Waves* (New York: W. H. Freeman, 1984), 94–96. By the 1820s, Fresnel believed that Young was receiving too much credit for reviving and extending the wave theory, and Worrall's account provides a methodological justification for that judgment. As Fresnel wrote in a letter to Young, "Monsieur, the merit of these discoveries, which belong to you exclusively by priority: also I have thought it useless to inform the public of those things which I discovered independently but after you; and if I speak of them to you, it is only to justify my paradoxical proposition, *that the apple would have come without the tree*" (96). Segrè comments with understatement that Fresnel is "too modest" in his letter. For additional support, see Jed Buchwald, *The Rise of the Wave Theory of Light: Optical Theory and Experiment in the Early Nineteenth Century* (Chicago: University of Chicago Press, 1989).

19 In 1818, Fresnel proposed a mathematical theory of diffraction in a French Academy of Sciences competition in opposition to corpuscular theory. Poisson showed that Fresnel's theory implied the appearance of a bright spot at the center of the shadow made by an illuminated disk. The prediction was thought to be an obvious refutation until it was checked experimentally and turned into a spectacular novel prediction.

20 Quoted in Musgrave, "Why Did Oxygen Supplant Phlogiston?" 191. Kirwan was a leading London chemist and defender of phlogiston.

21 Quoted in Aaron Ihde, *The Development of Modern Chemistry* (New York: Harper and Row, 1965), 61.

22 Musgrave's view of discovery in this context was provided earlier by Kuhn in 1962 in "The Historical Structure of Scientific Discovery," in *The Essential Tension: Selected Studies in Scientific Tradition and Change* (Chicago: University of Chicago Press, 1977), 167 ff. Musgrave uses a research programme analysis to arrive at the same conclusion. In this way, the research programme account provides a partial response to Kuhn's request for "a new vocabulary and new concepts for analyzing events like the discovery of oxygen" (171). See also *SSR*, 52 ff.

23 See Musgrave, "Why Did Oxygen Supplant Phlogiston?" 198; On numerous problems for the phlogiston programme, see also ibid., 196 n. 57.

24 Lavoisier created his notational system, published in 1787, with Louis Morveau. On Lavoisier's book as not containing fundamentally new content, see Ihde, *Development of Modern Chemistry,* 81.

25 In an 1857 paper titled in translation "The Kind of Motion Which We Call Heat," Rudolf Clausius related the differences between solid, liquid, and gaseous states of matter to molecular motion.

26 See Clark, "Atomism versus Thermodynamics," 42–44; Einstein is included as promoting this view. For a discussion of Clark's account, and the interaction of philosophy and science generally, see Margaret Morrison, "Scientific Conclusions and Philosophical Arguments," in *Scientific Practice: Theories and Stories of Doing Physics,* ed. Jed Buchwald (Chicago: University of Chicago Press, 1995).

27 For the kinetic and phenomenological heuristics, see Clark, "Atomism versus Thermodynamics," 42, 64.

28 See Stephen Brush, *The Kind of Motion We Call Heat: A History of the Kinetic Theory of Gases in the Nineteenth Century* (Amsterdam: North-Holland, 1976), 1: 78.

29 See Clark, "Atomism versus Thermodynamics," 44; on Duhem and Mach, see ibid., 91–92.

30 Quoted in Abraham Pais, *"Subtle Is the Lord . . .": The Science and Life of Albert Einstein* (New York: Oxford University Press, 1982), 82. On the "overdetermination" of Avogadro's number, see also ibid., 95. As Feyerabend points out, Einstein's predictions about Brownian motion were counterinductive, being in contradiction with the second law of thermodynamics as one of the most strongly supported empirical generalizations of nineteenth-century physics; see *Against Method* (London: New Left Books, 1975), 39–40.

31 Feyerabend argues that the separation of scientific, technological, theological, political-pragmatic (for instance, calendrical reform), and other influences in historiography of the Copernican revolution is especially artificial and misleading in *Science in a Free Society* (London: New Left Books, 1978), 41.

32 Thomas Kuhn, *The Copernican Revolution: Planetary Astronomy in the Development of Western Thought* (Cambridge, Mass.: Harvard University Press, 1957), 142–43. In addition, the "apparent economy of the Copernicus system, though it is a propaganda victory that the proponents of the new astronomy rarely failed to emphasize, is largely an illusion. . . . The seven-circle system . . . is a wonderfully economical system, but it does not work. It will not predict the position of planets with an accuracy comparable to that supplied by Ptolemy's system. . . . Copernicus, too, was forced to use minor epicycles and eccentrics. . . .

[H]is cumbersome sun-centered system gave results as accurate as Ptolemy's, but it did not give more accurate results" (168).

33 Ibid., 180.

34 Feyerabend, *Problems of Empiricism,* 225; see also 225 n. 43.

35 See chapter 8 below on the Michelson-Morley experiments.

36 Other physicists including Wilhelm Wien, Max Abraham, and Walter Kaufmann contributed to the electromagnetic programme, and well after 1905, textbooks and articles on the ether-electromagnetic approach continued to be published—for instance, on the behavior of the ether within and in the neighborhood of moving matter; see Russell McCormmach, "H. A. Lorentz and the Electromagnetic View of Nature," *Isis* 61 (1970): 459–97.

37 See Russell McCormmach, "Einstein, Lorentz, and the Electron Theory," *Historical Studies in the Physical Sciences* 2 (1970): 50.

38 Zahar, "Why Did Einstein's Programme Supersede Lorentz's?" 215.

39 See McCormmach, "Electromagnetic View," 468–69. This violation of Newton's third law provides another striking example of Lakatos's "progress on inconsistent foundations"; see below on Bohr's early quantum theory.

40 Zahar's arguments against the ad hoc character of Lorentz's electron theory falsify the views of many historians and Lorentz's contemporaries. Many understood the connection between Lorentz's molecular forces hypothesis and the Lorentz-Fitzgerald contraction, but also that it was a "plausibility" argument based on weak reasoning; a "dramatic" prediction of a Michelson null result it was not. See Arthur Miller, *Albert Einstein's Special Theory of Relativity: Emergence (1905) and Early Interpretation (1905–1911)* (Reading, Mass.: Addison-Wesley, 1981), 30.

41 See McCormmach, "Einstein, Lorentz, and the Electron Theory," 65.

42 On ending the central electromagnetic programme problem of relating matter and the ether, see Tetu Hirosige, "The Ether Problem, the Mechanistic Worldview, and the Origins of the Theory of Relativity," *Historical Studies in the Physical Sciences* 7 (1976): 41.

43 This key idea of Zahar's is a conclusion of Stanley Goldberg, *Understanding Relativity: Origin and Impact of a Scientific Revolution* (Boston: Birkhauser, 1984), 149. Goldberg cites Zahar's paper later in his book with the exaggeration that Zahar, Lakatos, and others use "logical reconstructions . . . as a substitute for evidence in historical arguments" (306). But underneath the disagreement there is a significant and mostly unrecognized convergence on historical-methodological fundamentals; a similar miscommuniciation with historians occurs in Lakatos's account of the Michelson-Morley experiments.

44 On Einstein, heuristic, and constructive models, see McCormmach, "Electromagnetic View," 489 ff.; and Hirosige, "The Ether Problem," 55. This key idea is not discussed by Zahar, perhaps because of the difficulties created for a research programme perspective. This lacuna and other weaknesses are discussed in Feyerabend, *Problems of Empiricism,* 226–28.

45 Hirosige, "The Ether Problem," 54–55.

46 Quoted in Feyerabend, *Problems of Empiricism,* 98. For a related discussion on Einstein's intellectual freedom, see Pais, "*Subtle Is the Lord . . . ,*" 13.

8 | Contradiction and Hindsight

1 On Popper's bitter responses to Lakatos, see Karl Popper, "Lakatos on the Equal Status of Newton's and Freud's Theories," in *The Philosophy of Karl Popper,* ed. Paul Schlipp, vol. 2 (LaSalle, Canada: Open Court, 1974).

2 On Popper's argument against Hegel, see Karl Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge* (New York: Harper Torchbooks, 1968): "If two contradictory statements are admitted any statement whatever must be admitted" (316). On the British Hegelian F. H. Bradley and relational contradictions, see Karl Popper, *The Open Society and Its Enemies,* 4th ed. (Princeton, N.J.: Princeton University Press, 1963), 1: 232; "Once a contradiction is admitted, all science must collapse" (ibid., 2: 38–39). Popper sometimes supported his ideas on philosophical logic using specious references to logical theories such as Alfred Tarski's formal semantics. See, for example, Karl Popper, *Objective Knowledge: An Evolutionary Approach* (Oxford: Oxford University Press, 1972), 330; and the discussion of ordered pairs in the look at Bradley mentioned above, in spite of Popper's advice that "among the real dangers to the progress of science is . . . a misplaced faith in formalization" (Popper, *Conjectures and Refutations,* 216).

3 Popper, *Conjectures and Refutations,* 322; see also *LSCD*, 91–92. Popper is restating that in formal logic, any statement is derivable from a contradiction—for instance, $(p \,\&\, {\sim}p) \rightarrow q$ is a tautology for all p and q. Lakatos's position simply implies that formal systems do not accurately reflect the empirical reason they are supposed to model.

4 And what gets regulated is Vernunft, or reason, whose limitations Kant wanted to demonstrate. The relevant perspective here on Kant is not his justification of synthetic a priori knowledge but the structural project of characterizing and coordinating sensibility, understanding, and reason.

5 Popper, *Conjectures and Refutations,* 38 n. 3.

6 G. W. F. Hegel, *Hegel's Logic,* pt. 1 of *Hegel's Encyclopedia,* trans. William Wallace (Oxford: Oxford University Press, 1975), 77–78, 117.

7 Kuhn's criticisms in *C&GK* (256–59) involve only the origins of Bohr's work and its role in the start of wave mechanics. Aside from these differences, Kuhn largely accepts Lakatos's account.

8 On setting out to prove and refute a mathematical conjecture, see also *P&R*, 37.

9 See Max Jammer, *The Conceptual Development of Quantum Mechanics* (New York: McGraw-Hill, 1966), 76.

10 Quoted in Jammer, *Conceptual Development,* 87. On the contrast between Bohr and Sommerfeld, see Andrew Whitaker, *Einstein, Bohr, and the Quantum Dilemma* (New York: Cambridge University Press, 1996), 121.

11 Ibid., 116.

12 Ibid.

13 We saw above in the Lakatos-Zahar account of Copernicus how Lakatos allowed novel fact to mean novel or unanticipated explanation of known fact.

14 For another account of the falsity of general scientific laws, with extensive examples, see Nancy Cartwright, *How the Laws of Science Lie* (New York: Oxford University Press, 1983). Where Lakatos provides a diachronic description of how a

hard core gets articulated through the protective belt of testable theories, Cartwright gives a synchronic account of high-level laws, like Maxwell's equations, and the models needed to apply them to real problems. Thus, as Cartwright says, there is a "trade-off" realized between "truth and explanatory power" (59).

15 Alfred Schmidt, *The Concept of Nature in Marx,* trans. Ben Fowkes (London: New Left Books, 1971), 184.

16 Hegel, *Hegel's Logic,* 77; Hegel's extended account is found in his *Philosophy of Nature.* On this topic and its Kantian-objectivist themes, see also Lucio Colletti, *Marxism and Hegel,* trans. Lawrence Garner (London: Verso, 1973). For an account of Engels's dialectics of nature, see Leszek Kolakowski, *Main Currents of Marxism,* trans. P. S. Falla (New York: Oxford University Press, 1981), 1: chapter 15. Colletti sums up Engels's *Dialectics of Nature* as "a work 90 per cent of which is hopelessly compromised by an ingenuous and romantic *Naturphilosophie,* contaminated by crudely positivist and evolutionist themes" ("A Political and Philosophical Interview," in *Western Marxism: A Critical Reader,* ed. New Left Review [London: Verso, 1978], 329).

17 Immanuel Kant, *Critique of Pure Reason,* trans. Norman Kemp Smith (New York: St. Martin's Press, 1965), A302/B359 (302).

18 See chapter 9, note 13, below; and text.

19 G. W. F. Hegel, *Hegel's Science of Logic,* trans. A. V. Miller (New York: Humanities Press, 1976), 440. In an 1827 discussion with his lifelong friend Goethe, Hegel said that dialectical method was nothing but the innate capacity for working through contradictions, to which Goethe replied that he hoped this capacity would not be misused and "employed to make the false true and the true false." Hegel claimed it could happen, "but only with people who are mentally diseased." Goethe responded that he was "also certain that many a dialectic disease would find a wholesome remedy in the study of nature," emphasizing that nature "throws off as incapable everyone who does not proceed purely and honestly with the treatment and observation of his subject" (Johann Wolfgang von Goethe, *Conversations with Eckermann, 1823–1832,* trans. John Oxenford [San Francisco: North Point Press, 1984], 198). On the influence of Hegel and *Naturphilosophie* in Prussia, and the mutual disrespect between physicists and nature philosophers, see Christa Jungnickel and Russell McCormmach, *The Intellectual Mastery of Nature: Theoretical Physics from Ohm to Einstein* (Chicago: University of Chicago Press, 1986), 1: 27.

20 See also Hegel, *Hegel's Logic,* 77, 117; and, for example, the section on Kant in G. W. F. Hegel, *Lectures on the History of Philosophy,* trans. E. S. Haldane and F. H. Simson (New York: Humanities Press, 1974), 3: 443 ff.

21 Kant, *Critique of Pure Reason,* A644/B672.

22 Ian Hacking, "Imre Lakatos' Philosophy of Science," *British Journal for the Philosophy of Science* 30 (1979): 383 ff.

23 "In Marx, as in Kant, the form and the matter of the phenomenal world can be separated *in abstracto* but not in reality. It is ultimately meaningful to refer to the Kantian problem of constitution when discussing Marx's dialectic, because Marx followed Kant in holding that form and matter are external to one another, despite the great difference between their views on the way the two elements interact. . . .

Marx adopted an intermediate position between Kant and Hegel, which can only be fixed with difficulty. His materialist critique of Hegel's identity of Subject and Object led him back to Kant, although again this did not mean that being, in its non-identity with thought, appeared as an unknowable 'thing-in-itself.' . . . Marx both retained Kant's thesis of the non-identity of Subject and Object and adopted the post-Kantian view, no longer exclusive of history, that Subject and Object entered into changing configurations" (Schmidt, *Concept of Nature in Marx,* 120–21). For a similar "historicized neo-Kantianism," see Peter Galison, *Image and Logic: A Materialist Culture of Microphysics* (Chicago: University of Chicago Press, 1997), 840; unsurprisingly, Galison uses the word "materialist" in his subtitle.

24 For Popper on Michelson-Morley, see *LSCD*, 108. While Lakatos sides with those who deny significant knowledge on Einstein's part, his account of the Michelson-Morley experiments does not rely on that. Most of my account of the Michelson-Morley experiments is from Loyd Swenson, *The Ethereal Aether: A History of the Michelson-Morley-Miller Aether-Drift Experiments, 1880–1930* (Austin: University of Texas Press, 1972). Michelson's 1881 and Michelson and Morley's 1887 papers are reproduced in Swenson's book as appendices.

25 Swenson, *Ethereal Aether,* 257. Ian Hacking criticizes Lakatos's account of the Michelson-Morley experiments in *Representing and Intervening* (New York: Cambridge University Press, 1983), 253–61. Hacking takes Lakatos to task for the falsifications of his narratives generally—for example, Lakatos is said to be "lying": "no point was being served. . . . [Lakatos] made a couple of silly remarks . . . where he asserts something as historical fact in the text, but retracts it in the footnotes . . . [and so] the historical reader is properly irritated" (125). My view, however, is that Lakatos was purposely maintaining a "strict" research programme approach to demonstrate the theory-ladenness of any historiographical technique. Lakatos thought that Kuhn and others never grasped his point about the codependence of historical narrative and methodology (see *MSRP*, 120 n. 1, 192), and that is Hacking's blind spot about Lakatos, too. Given Lakatos's goals, except for a mistaken dating of Michelson's last experiment as 1935 instead of 1928, his philosophical-historical reconstruction should not be rewritten. The discussion in the text of Michelson's role as an experimenter and not a theorist also reflects Hacking's criticism of Lakatos's bias toward theoretical science.

26 Swenson, *Ethereal Aether,* 270.

27 Ibid., 281.

28 Ibid., 118.

29 On Shankland, see Swenson, *Ethereal Aether,* 242 ff. Ironically and sadly, when Michelson visited Helmholtz in Berlin in 1880, the latter had warned that temperature fluctuations would be difficult to control, and these were indeed Miller's undoing on Mount Wilson.

30 Quoted in Swenson, *Ethereal Aether,* 207.

31 Swenson, *Ethereal Aether,* 176 ff.

32 Albert Einstein, "On the Electrodynamics of Moving Bodies," in Hendrik Lorentz et al., *The Principle of Relativity,* trans. W. Perrett and G. B. Jeffery (New York: Dover, 1952); 37. For another account also minimizing the importance of Michel-

son's work for Einstein, see Gerald Holton, "Einstein, Michelson, and the 'Crucial' Experiment," in *Thematic Origins of Scientific Thought: Kepler to Einstein,* rev. ed. (Cambridge, Mass.: Harvard University Press, 1973). This chapter of Holton's book appeared originally in 1969. Perhaps confusion surrounding Lakatos's rational reconstructions made his most significant historical theses opaque in spite of agreement on fundamentals and shared opposition to some received views of the time.

33 See Swenson, *Ethereal Aether,* 188–89.

34 G. W. F. Hegel, *Hegel's Philosophy of Right,* trans. T. M. Knox (New York: Oxford University Press, 1979), 12.

35 *FM*, 65; *MSE*, 209; and *FM*, 99.

36 "One may illustrate this using the Bohrian program. Bohr, in 1913, may not have even thought of the possibility of electron spin. He had more than enough on his hands without the spin. Nevertheless, the historian, describing with hindsight the Bohrian program, should include electron spin in it, since electron spin fits naturally in the original outline of the program. Bohr might have referred to it in 1913. Why Bohr did not do so, is an interesting problem which deserves to be indicated in a footnote" (*HS*, 119–20). And Lakatos fulfills this promise with a footnote.

37 For an example of normative advice, see Lakatos on the Bohr-Kramers-Slater theory in *FM*, 84.

9 | Reason in History

1 By joining fallibility to historicism and interpretation, two types of "falsification" now come into play. One is to recognize the anomalies or contradictions of science or history as overcome through an improved theory. The other is to distort history through the lens of some historiographical theory.

2 See also *MSE*, 178 n. 3: "Because of the imperfection of the scientists, some of the actual history is a caricature of its rational reconstruction; because of the imperfection of the methodologists, some methodologies are caricatures of actual history. (And, one may add, because of the imperfection of historians, some histories of science are caricatures both of actual history and of its rational reconstruction.)" Note that Lakatos does not claim anywhere what history is *not* some type of caricature.

3 See chapter 10, note 5, and text below.

4 For an account bearing comparison to Lakatos's, see Hayden White, *Tropics of Discourse: Essays in Cultural Criticism* (Baltimore, Md.: Johns Hopkins University Press, 1978). For example: "The distinction between proper history and metahistory tends to dissolve into a matter of emphasis" (62); "The governing metaphor of an historical account could be treated as a *heuristic rule which self-consciously eliminates certain kinds of data from consideration as evidence*" (46, emphasis in original); or, "Many historians continue to treat their 'facts' as though they were 'given' and refuse to recognize, unlike most scientists, that they are not so much found as constructed by the kinds of questions which the investigator asks of the phenomena before him" (43). White's *Metahistory: The Histori-*

cal Imagination in Nineteenth-Century Europe (Baltimore, Md.: Johns Hopkins University Press, 1973) provides an account of Marx's historiography, but without addressing Lukács's perspective or Western Marxist views generally, which were becoming better known in the English-speaking world around the time that *Metahistory* was published. Relevant to the many considerations on language related to Lakatos's work is White's comment in *Metahistory* that "it is no accident that the principal philosophers of history were also (or have lately been discovered to have been) quintessentially philosophers of language" (xi).

5 "The concepts of 'progressive' and 'degenerating' problemshifts, the idea of [the] proliferation of theories can be generalized to any sort of rational discussion and thus serve as tools for a general theory of criticism. . . . [T]he methodology of scientific research programmes may be applied not only to norm-impregnated historical knowledge but to any normative knowledge, including even ethics and aesthetics" (*FM*, 92 n. 3; *HS*, 133 n. 4).

6 On "quasi-empirical," and the interesting suggestion that "the [methodological] 'quasi-Euclidean' versus 'quasi-empirical' distinction is more useful than the [epistemological] '*a priori*' versus '*a posteriori*' distinction," see *HS*, 136 n. 2.

7 Hegel also discussed the internal/external distinction in the introduction to his *Lectures on the History of Philosophy,* in *Hegel's Idea of Philosophy,* ed. and trans. Quentin Lauer (New York: Fordham University Press, 1974), 93. See also Ferenc Fehér, "Lukács in Weimar," in *Lukács Reappraised,* ed. Ágnes Heller (New York: Columbia University Press, 1983): "One of Lukács' fundamental character traits returns here, a feature that originated from his pre-Marxist youth: his hatred for psychology . . . because the 'soul' . . . is a result of free action. . . . A methodology adequate to cope with this negative attitude is pursued by Lukács consistently up to the end of his active life. With him writers have no biographies" (91).

8 On Robert Merton and the social-functionality of priority disputes, for example, see *HS*, 116 n. 3.

9 The idea of false consciousness appears in *RH* (see 72–93), and is tied both to the "cunning of reason" (the net result of people's actions may have a rational structure in spite of quite different individual motivations) and Hegel's "world-historical individuals," who are able to correctly grasp the direction of their culture's movement in order to first formulate or act on it. As we will see below, a version of the latter elite occurs in the methodology of scientific research programs, while the cunning of reason does not. Lakatos believes in no "true consciousness," in that any methodology "can be 'falsified' for the simple reason that no set of human judgments is completely rational and thus no rational reconstruction can ever coincide with actual history" (*HS*, 131). For a similar idea in the context of the Hungarian Revolution, see Tamás Aczél, "The Mythology of True Consciousness," in *Ten Years After: A Commemoration of the Tenth Anniversary of the Hungarian Revolution* (London: Macgibbon and Kee, 1966), 190–203.

10 Quoted in Popper, *The Open Society and Its Enemies,* 4th ed. (Princeton, N.J.: Princeton University Press, 1963), 1: 186. Lakatos's nondeterministic historicism combined with his critical apparatus provides a fairly complete rebuttal to many of the wrongheaded ideas about Hegel and history found in *The Open Society.*

Note that George Soros, the Hungarian billionaire financier, was greatly influenced by Popper's ideas on criticism and the open society—Soros contributed significantly through his foundation to opposition movements before the fall of the Berlin Wall—but also has an explicitly Hegelian understanding of economics and financial markets; see his *Alchemy of Finance* (New York: Simon and Schuster, 1987).

11 The assessment of degeneration is often problematic since one never knows how long to wait for an improvement. Lakatos thought that the particle physics of the early 1970s was degenerating (see *HS*, 137 n. 2), but just after his death in 1974, and then into the 1980s, the discovery of quarks and other types of particles enabled the rapid development and confirmation of the "standard model." See Frank Close et al., *The Particle Explosion* (New York: Oxford University Press, 1987), 176; on the uninformativeness of research programmes for the judgment of cancer research, see Ian Hacking, *Representing and Intervening* (New York: Cambridge University Press, 1983), 118. But important judgments of progress or degeneration are being made, such as the judgment that research on the determination of cancer effects due to electromagnetic field exposure showed no progress, or that global warming is occurring. Supposedly, these judgments are independent of the costs and benefits of acting on them, but what happens is that a weak methodological appraisal is subtly amplified because of the high stakes involved. Hence, the methodology of scientific research programmes can still be used to help calibrate the extent to which that is occurring and its influence on public policy. On problems in the assessment of uncertainty in public policy using a philosophy of science approach, see Silvio Funtowicz and J. R. Ravetz, *Uncertainty and Quality in Science for Policy* (Dordrecht, Netherlands: Kluwer, 1990); on climate change and electromagnetic field research progress, see my "Taming Chance: Risk and the Quantification of Uncertainty," *Policy Sciences* 29 (1996): 1–27. For a detailed examination with weak support for Lakatos's normative use of research programme comparisons, see Alan Musgrave, "Method or Madness: Can the Methodology of Scientific Research Programmes Be Rescued from Epistemological Anarchism?" in *Essays in Memory of Imre Lakatos,* ed. Robert Cohen et al. (Boston: D. Reidel, 1976). For a direct challenge with historical examples to the logical coherence of Lakatos's ideas for historiographical comparisons, see Husain Sarkar, *A Theory of Method* (Berkeley: University of California Press, 1983).

12 Gerald Holton, *Thematic Origins of Scientific Thought: Kepler to Einstein,* rev. ed. (Cambridge, Mass.: Harvard University Press, 1973), 459, with Holton quoting Lakatos from "Understanding Toulmin" (*MSE*, chapter 11).

13 Paul Feyerabend, *Problems of Empiricism: Philosophical Papers, Volume 2* (New York: Cambridge University Press, 1981), 34. For a similar charge of schizophrenia, see *MSRP*, 212.

14 Feyerabend, *Problems of Empiricism,* 34.

15 Ibid., 46, with Feyerabend quoting from Vasco Ronchi, *Historie de la Lumière.*

16 Lakatos quoted in Yehuda Elkana, ed. *The Interaction between Science and Philosophy* (Atlantic Highlands, N.J.: Humanities Press, 1974), 292. On historical caricature, see also the conclusion in *FM*, 92.

17 Since the debates of the 1960s and 1970s, the philosophy of science has been partially eclipsed by many types of science studies. Lakatos figures little as an inspira-

tion for this work, which has been greatly influenced by Kuhn, Feyerabend, and others including Michel Foucault. One difference from Lakatos's work is a diminished interest in a critical or normative analysis of the production of ideas in favor of descriptions of scientific practices. But Lakatos's historiographical point that judgments are nonetheless always made about what counts as a valid inference, or even a scientific act, is still valid. Nazi accusations of "Jewish science," or even the petty anti-Semitism faced by the physicist Richard Feynman among others, are, I assume, still seen as external, as is the influence of industrial science sponsors in environmental, pharmaceutical, or tobacco research. Without some conception of what counts as science, the choice of what to include in any history, as Lakatos emphasizes, would seem again to contain some arbitrariness, and it is unclear how one formulates or justifies a social-critical perspective. At the same time, while Lakatos emphasized that external history was just as necessary as internal history, he provided zero guidance on how to integrate them. Much in contemporary science studies, in contrast, does just that, through conceptions of: "constraints," linguistic "trading zones," and experimental method (Peter Galison, "Context and Constraints," in *Scientific Practice: Theories and Stories of Doing Physics,* ed. Jed Buchwald [Chicago: University of Chicago Press, 1995], and *Image and Logic*); the many-layered and tacit choices of research execution (Andrew Pickering, *The Mangle of Practice* [Chicago: University of Chicago Press, 1995]); the quasi-autonomy of experimental work (Hacking, *Representing and Intervening*); the coercive role of knowledge production and use, for example in biomedical characterizations of deviancy (Stephen J. Gould, *The Mismeasure of Man* [New York: W. W. Norton, 1981]); the productive roles for institutional or cultural settings in promoting research or scientific methods (Steven Shapin, *A Social History of Truth: Civility and Science in Seventeenth-Century England* [Chicago: University of Chicago Press, 1994]); and the positive importance in contemporary science for managerial, engineering, and even entrepreneurial roles (Paul Rabinow, *Making PCR: A Story of Biotechnology* [Chicago: University of Chicago Press, 1996]). These are all not obviously helped through Lakatos's historiographics and especially his overly rigid internal-external distinction.

10 | A Changing Logic

1 At the time of his death, Lakatos had planned a book to be titled *The Changing Logic of Scientific Discovery,* an obvious challenge to Popper's static *LSCD* (*FM*, 8 n.*). While Lakatos never claimed that his skeptical construction defined a "changing logic," the affinities to the background in Hegel and Lukács are strong enough to warrant exploring the interpretation. This reconstruction of Lakatos's changing logic should not be ascribed to him.

2 Lakatos presents several historiographical models in *HS*: historiographical inductivism, conventionalism, Popperian methodological falsificationism (all of which are falsified by history), and then the methodology of scientific research programmes. These again form the skeleton of the methodology of scientific research programmes itself, and are taken by Lakatos as defining successively progressive historiographical research programmes.

3 The idea was also developed around the same time by Karl Korsch. For an account

of Lukács's ideas during this time and *H&CC*, see Andrew Arato and Paul Breines, *The Young Lukács and the Origins of Western Marxism* (New York: Continuum/Seabury, 1979).

4 This is a bowdlerized version of *H&CC*, 47–48, starting with the last paragraph on 47. On "historical decline," see *P&R*, 98 n. 2, 154.

5 See Rüdiger Bubner, *Modern German Philosophy*, trans. Eric Matthews (New York: Cambridge University Press, 1981), 165.

6 On Hegel, Kant, and the criterion, see Jürgen Habermas, *Knowledge and Human Interests*, trans. Jeremy Shapiro (Boston: Beacon Press, 1971), chapter 1; and John Smith, "Hegel's Critique of Kant," *Review of Metaphysics* 26 (1973): 438–60.

7 G. W. F. Hegel, *Hegel's Logic*, pt. 1 of *Hegel's Encyclopedia*, trans. William Wallace (Oxford: Oxford University Press, 1975), section 41 (66).

8 Ibid.

9 Ibid.

10 Ibid.

11 Quoted in Ronald Gray, *Goethe: A Critical Introduction* (New York: Cambridge University Press, 1967), 200. Schlegel is remarking on the numerous changes in narrative framing occurring in the text.

12 G. W. F. Hegel, *Hegel's Science of Logic*, trans. A. V. Miller (New York: Humanities Press, 1976), 43, with "Concept" for Hegel's "Begriff."

13 Ibid., 54.

14 Ibid., 53.

15 Ibid., 842.

16 Ibid., 843.

17 Quoted in Roman Rosdolsky, *The Making of Marx's "Capital,"* trans. Pete Burgess (London: Pluto Press, 1977), 4 n. 21.

18 Quoted from Marx's *Grundrisse* in ibid., 44.

19 Habermas, *Knowledge and Human Interests*, 42.

20 Ibid., 63.

21 Ibid.

22 Lucio Colletti, "Marxism and the Dialectic," *New Left Review* 93 (1975): 21–22. Colletti also asks, "Can a purely scientific theory contain within itself a discourse on alienation? The problem has not yet been resolved" ("A Political and Philosophical Interview," in *Western Marxism: A Critical Reader*, ed. New Left Review [London: Verso, 1978], 336). The text provides a sketch of an answer.

11 | Classical Political Economy as a Research Programme

1 Adam Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations* (New York: Modern Library, 1937), 58.

2 Smith, *Wealth of Nations*, 33.

3 Before Smith, John Locke suggested that labor is the main source of value, but primarily in the context of use value and without distinguishing that from exchange value. Marx himself introduced the term "classical economists" to denote the school beginning with the "mercantilists" William Petty in England and Pierre Boisguillebert in France, to David Ricardo and Leonard Sismondi.

4 One of Marx's key contributions was to see the need for explaining the reproduction of conditions of production themselves as part of this scheme.

5 On this shift to a more internal and autonomous development of the labor theory, see Ronald Meek, *Studies in the Labour Theory of Value,* 2d ed. (London: Lawrence and Wishart, 1973), 82 ff. See also George Lichtheim, *Marxism,* 2d ed. (New York: Praeger, 1969), 170.

6 My account is consistent with projects to reconstruct Marx's social theory without the labor theory of value, for example, as in Moishe Postone, *Time, Labor, and Social Domination: A Reinterpretation of Marx's Critical Theory* (New York: Cambridge University Press, 1996). That is, Marx's immanent critique of political economy and his historicization of economic categories are compatible with his role as an economic scientist, and the problems of the latter can be acknowledged without rejecting the usefulness of much in Marx's social theory, such as the analysis of the roles for labor and time in capitalist economies. A question is whether one should continue the analysis to a more empirical level of explaining market phenomena, and how that then might be done. Clearly, the labor theory of value provides little or no help today, and Marx's attempt at solving the transformation problem needs to be recognized as a failure. Postone (135 ff.) is ambiguous on the present-day role of the transformation problem in a reconstruction of Marx's ideas; he is also not critical enough of Marx's lack of success. But Marx's limits can be recognized, and the transformation problem abandoned, without vitiating the integrity or coherence of his social-economic categories as a whole, and as explored by Postone. Filling in *whatever way* the empirical modeling "gap" left by the removal of the "problem" of transforming values into prices should be seen as a desirable challenge to be embraced, and neither argued away nor ignored. For additional discussion, see Verso–NLB, eds., *The Value Controversy* (London: Verso, 1981).

7 Another crucial component of Marx's economic research is his theory of business cycles developed in volume 2 of *Capital.* Business cycles were a mysterious novelty in the mid–nineteenth century, with Marx being among the first to recognize their importance and many of their principal features. They are not included for discussion here because the relation of Marx's account to the labor theory of value is at best indirect; moreover, the labor theory is primarily an equilibrium theory of prices, while business cycles inherently entail dynamic problems of disequilibrium.

8 See Maurice Dobb, *Theories of Value and Distribution since Adam Smith* (New York: Cambridge University Press, 1973), 70.

9 David Ricardo, *Principles of Political Economy and Taxation* (Buffalo, N.Y.: Prometheus, 1996), 17. A fair comparison to facilitate the interpretation of these theories is that "value" is an abstraction and a theoretical term like modern economic "utility." It is not descriptive of anything directly observed in the marketplace but is a mixture of tacit empirical generalizations and simplifying assumptions; its role thus varies considerably across theorists. Robert Paul Wolff argues, like the economist Joan Robinson several decades ago, that Marx himself effectively hypostasized his value and labor abstractions, attributing to them a far greater degree of realism than warranted (see *Understanding Marx: A Recon-*

struction and Critique of "Capital" (Princeton, N.J.: Princeton University Press, 1984), 115, 142 ff. Karl Popper in *The Open Society and Its Enemies,* 4th ed. (Princeton, N.J.: Princeton University Press, 1963) similarly says that Marx "made a great effort to destroy this mystical character of objective 'value,' but he did not succeed" (2: 177).

10 Ricardo, *Principles,* 19, 29.

11 Marx compared Smith and Ricardo's approaches to the length of an imaginary line divided into three unequal segments. Smith saw the three segments as being drawn first and then being pasted together to form the longer line, and whose length varied with the lengths of the segments that are unconstrained. Ricardo and Marx saw the longer line with a fixed length being divided into three segments whose lengths and total length are constrained. See Karl Marx, *Capital,* ed. Friedrich Engels, trans. Samuel Moore and Edward Aveling, 3d ed. (New York: International Publishers, 1967), 2: 386.

12 "The principle that the quantity of labour bestowed on the production of commodities regulates their relative value, considerably modified by the employment of machinery and other fixed and durable capital" (Ricardo, *Principles,* 30); and "The principle that value does not vary with the rise or fall of wages, modified also by the unequal durability of capital, and by the unequal rapidity with which it is returned to its employer" (ibid., 35). Prior to Piero Sraffa's edition of Ricardo's *Works and Correspondence* it was thought that Ricardo, like Smith, had "retreated" from the labor theory, while his view now appears to be more nuanced than a simple accept/reject dualism.

13 Between 1820 and 1840, a group of "left-wing" Ricardians (including Piercy Ravenstone, John Gray, William Thompson, Thomas Hodgskin, and J. F. Bray) drew on socialist theory and Ricardo's labor doctrine. Technically "pre-Marxians," their ability to develop theoretical economics is not comparable to that of Ricardo or Marx. See George Lichtheim, *The Origins of Socialism* (New York: Praeger, 1969), chapter 8.

14 Wolff, *Understanding Marx,* 91. For evidence on the methodological structure of volumes 1 and 3, see 92 ff.

15 Quoted from a letter to Ludwig Kugelmann in Meek, *Studies in the Labour Theory of Value,* 153.

16 On turning Ricardo's so-called exceptions into "law," see *TSV*, 3: 176–77; and on similar explanations related to economic crises, see *TSV*, 2: 499, 512, 519.

17 "This section, which Ricardo added to his third edition, bears witness to his *honesty* which so essentially distinguishes him from the vulgar economists" (*TSV*, 3: 555).

18 See *P&R*, 46 n. 1, 87 n. 1, 134–36.

19 For a discussion of this old problem, see Jürgen Habermas, *Knowledge and Human Interests,* trans. Jeremy Shapiro (Boston: Beacon Press, 1971), chapter 3.

20 Analogous to "necessity," Marx's use of "phenomena" and "essence," sometimes mocked as metaphysical categories, here too have a pure methodological content; see *TSV*, 2: 500–501. For a standard "metaphysical" interpretation, see Mark Blaug, *Economic Theory in Retrospect,* 3d ed. (New York: Cambridge University Press, 1978), 281.

21 Lichtheim, *Marxism,* 172, 175 (first two quotations); and Blaug, *Economic Theory*

in Retrospect, 300. While providing several criticisms, and ultimately seeing it as dispensable, Popper offers a surprisingly favorable account of Marx's theory of surplus value in *The Open Society:* "A theoretical success of the first order. . . . [E]ven if Marx's analysis was defective, his effort to explain the phenomenon of 'exploitation' deserves the greatest respect" (2: 173, 178). Marx's transformation of values into prices is in *Capital,* vol. 3, chap. 9, "Formation of a General Rate of Profit (Average Rate of Profit) and Transformation of the Values of Commodities into Prices of Production."

22 Marx's role in rehabilitating the labor theory was unique, with post-Ricardian political economy exclusive of Marx not adhering closely to the labor theory, as argued in Mark Blaug, "Kuhn versus Lakatos *or* Paradigms versus Research Programs in the History of Economics," in *Method and Appraisal in Economics,* ed. Spiro Latsis (New York: Cambridge University Press, 1976), 165–66. Blaug "will also resist the temptation to apply *MSRP* to Marxian economics, which began badly to 'degenerate' in the first decade of the [twentieth] century" (67 n. 61), but Marxian economics begins with Marx and his research program began before him. In his *Methodological Appraisal of Marxian Economics* (Amsterdam: North-Holland, 1980), Blaug claims to apply the methodology of scientific research programmes to Marxian economics, yet he focuses only on the twentieth century and does not identify a hard core, heuristics, or progression of theories. His methodological ideas are mostly Popperian, some of which are not held by Lakatos—for instance, identifying conditions of refutation in advance. For other discussions of Lakatos, Popper, and economic methodology, see Roger Backhouse, *New Directions in Economic Methodology: From Lakatos to Empirical Philosophy of Science* (New York: Routledge, 1998).

23 Marx, *Capital,* 1: 309.

24 The mathematical economist Michio Morishima argues that Marx faced difficult mathematical problems dealt with by mathematicians only decades later—such as the existence of eigenvalues for nonnegative square matrices needed in the matrix representation of input-output systems, and answered by the 1907 Frobenius-Perron theorem. See "Marx in the Light of Modern Economic Theory," *Econometrica* 42 (1974): 611–32.

25 Marx, *Capital,* 3: 198.

26 Ibid., 3: 173.

27 Blaug, *Economic Theory in Retrospect,* 224–25.

28 Marx's failure to transform inputs and the problem of his two so-called conservation requirements was first analyzed by Bortkiewicz, and later in simpler forms by Joseph Winteritz in 1948 and Francis Seton in 1957; see Meek, *Studies in the Labour Theory of Value,* xxiii, 95–96.

29 On "one-sector" economies for which Marx's assumptions hold, and the relations between the simple labor theory and Marx's two conditions, see Wolff, *Understanding Marx,* chapter 6.

30 On Marx's understanding of the need to transform inputs as well as outputs, see Ian Steedman, *Marx after Sraffa* (London: Verso, 1977), 31ff.

31 Meek, *Studies in the Labour Theory of Value,* 314.

32 Ibid.

33 Joseph Gillman, *The Falling Rate of Profit* (London: Denis Dobson, 1957), 59.

34 Blaug, *Economic Theory in Retrospect,* 261, commenting on Gillman's study.
35 Shlomo Avineri, *The Social and Political Thought of Karl Marx* (Cambridge, U.K.: Cambridge University Press, 1971), 256.
36 Leszek Kolakowski, *Main Currents of Marx,* trans. P. S. Falla (New York: Oxford University Press, 1981), 1: 329.
37 See Steedman, *Marx after Sraffa,* 29ff., 52, 116ff.
38 On further approaches, see E. K. Hunt and Mark Glick, "Transformation Problem," in *The New Palgrave Marxian Economics,* ed. John Eatwell et al. (New York: W. W. Norton, 1987), 360.
39 Dobb, *Theories of Value and Distribution,* 265.
40 Samuelson's withering 1957 judgment of Marx is that he is "a minor post-Ricardian" and an "autodidact" (quoted in Wolff, *Understanding Marx,* 4). To drive the point home that Marx's labor theory formulation is thoroughly misconceived, Steedman argues that Marx's linear value definition can go negative; see *Marx after Sraffa,* chapter 11. Sraffa's approach also implies a limited theoretical model, but at least it is internally consistent.
41 See John Roemer, "Marxian Value Analysis," in *The New Palgrave Marxian Economics,* ed. John Eatwell et al. (New York: W. W. Norton, 1987), 262.

12 | Hungary 1956 and the Inverted World

1 See Ferenc Váli, *Rift and Revolt in Hungary* (Cambridge, Mass.: Harvard University Press, 1969), 206. Other accounts include György Litván, ed., *The Hungarian Revolution of 1956: Reform, Revolt, and Repression, 1953–1963,* trans. János Bak and Lyman Legters (New York: Longman, 1996); Béla Király et al., eds., *The First War between Socialist States: The Hungarian Revolution of 1956 and Its Impact* (New York: Brooklyn College Press, 1984); and Charles Gati, *Hungary and the Soviet Bloc* (Durham, N.C.: Duke University Press, 1986). On Stalinization and social paranoia, see also Paul Ignotus, *Hungary* (New York: Praeger, 1972), chapter 10; and R. J. Crampton, *Eastern Europe in the Twentieth Century* (New York: Routledge, 1994), 267.
2 See George Hodos, *Show Trials: Stalinist Purges in Eastern Europe, 1948–1954* (New York: Praeger, 1984), 5–7.
3 See *RM*, 251. Farkas junior was said to have read novelist Octave Mirbeau's *Le Jardin des Supplices* for advice on torture.
4 See David King, *The Commissar Vanishes: The Falsification of Photographs and Art in Stalin's Russia* (New York: Henry Holt, 1997).
5 Władysław Gomułka to the Central Committee of the Polish United Workers Party, 21 October 1956, quoted in Váli, *Rift and Revolt,* 37. Marx apparently coined the phrase "personality cult"; see the 1877 letter to W. Blos in which he described how he and Engels share an "aversion to any personality cult" (in *The Marx-Engels Reader,* ed. Robert Tucker, 2d ed. [New York: W. W. Norton, 1978], 521). In his secret speech denouncing Stalin to the twentieth congress following Stalin's death, Khrushchev cited the letter to support his claim that the personality cult was foreign to Marxism.
6 On Party structure and elitism, see Béla Szász, *Volunteers for the Gallows: Anatomy of a Show-Trial,* trans. Kathleen Szász (New York: W. W. Norton, 1971), 67ff.

7 Szász, *Volunteers for the Gallows,* 221. See also Hodos, *Show Trials:* "The starting point was always the autobiography of the arrested person, his own account of his political life. Only then could the interrogators begin a political reinterpretation of that life, achieved with rubber truncheons, rifle butts, electric shocks, sleeplessness, hunger, and cold—a mixture of the most advanced and archaically barbaric methods of physical and psychological tortures. The interrogators suggested 'reinterpretations' of previous arrests for underground communists' activities so that they became informer services for the fascist police. Personal or professional contacts with the co-accused became plotting with and spying for the agents of imperialism. Step by step, each version was transformed by torture into a new and uglier one until every moment of the accused's life was reinterpreted into an abominable crime" (46).

8 "Literature must become a part of the general cause of the proletariat, 'a small cog and a small screw' . . . a mechanism set in motion by the entire conscious vanguard of the whole working class" (Lenin quoted in *RM*, epigraph).

9 Lukács, the "short, ugly, pock-marked, big-eared old gentleman," sounds like a Hungarian-Marxist Socrates, whose "eyes frequently sparkled with childish pleasure and excitement, usually when he was explaining something. . . . He loved to explain. The listener often felt that this strange, kind, and polite man considered the aim and purpose of his life to be the explanation itself, not the problem he was struggling so hard to explain. . . . [If] one asks why so many writers embraced the cause of Communism in 1949, the reason lies, perhaps not exclusively but certainly to a large extent, in the fact that the Communist cause was expounded by the thin-fingered, wrinkled hands of a very intelligent, very erudite old man named György Lukács. . . . Almost all young people were disciples of Lukács in those days" (*RM*, 59, 67, 72).

10 Lukács "had learned tactics were all important. . . . [O]ne had to close one's eyes to horror and compromise, both in large and small matters, particularly if one wanted to survive and, more important, to go on working" (*RM*, 63).

11 Litván, *Hungarian Revolution,* 24–25.

12 See *RM*, 244 ff.

13 Another András Hegedűs was active in the Petőfi Circle and inner-Party opposition. See Litván, *Hungarian Revolution,* 200.

14 Lukács did not sign, apparently not because he disagreed but because he had learned to be careful with the Party. Vis-à-vis Madach's *Tragedy of Man,* Lucifer attacks the Lord, who replies, "I am to be revered, not criticized." The line would bring a storm of applause with the recognition of the exact character of the Rákosi regime (see *RM*, 353).

15 The passage on "non-Marxist" and so on is Gyula Háy's.

16 Georg Konrád, *The Loser,* trans. Ivan Sanders (New York: Harcourt Brace Jovanovich, 1982), 203.

17 Quoted in Litván, *Hungarian Revolution,* 141.

18 Litván, *Hungarian Revolution,* 143 ff.

19 Free Radio Kossuth, Budapest, November 4, 1956, 7:56 A.M., quoted in Melvin Lasky, ed., *The Hungarian Revolution: A White Book* (New York: Praeger, 1957), 228. The speakers are Gyula and Éva Háy.

20 Much of the following on Lakatos's life is told by Jancis Long in her superb

"Lakatos in Hungary," *Philosophy of the Social Sciences* 28 (1998): 245–311. I have also used Lee Congdon, "Possessed: Imre Lakatos' Road to 1956," *Contemporary European History* 6 (1997): 279–94. Some material is also from the biographical sketch by John Worrall, "Imre Lakatos (1922–1974): Philosopher of Mathematics and Philosopher of Science," in *Essays in Memory of Imre Lakatos,* ed. Robert Cohen et al. (Boston: D. Reidel, 1976), and discussions with people who knew Lakatos. Considerable uncertainty is associated with many details of Lakatos's life in Hungary. Congdon may overestimate Lakatos's role in the events he discusses due to the lack of knowledge of others involved. Lakatos's association with Révai may also have not have been especially strong, though there clearly were significant interactions.

21 See László Ropolyi, "Lukács and Lakatos," in *The Philosophy of Imre Lakatos: Its Roots, Content, and Limitations,* ed. Michael Stöltzner et al. (Dordrecht, Netherlands: Kluwer, forthcoming). In 1937, photographs of Lakatos/Lipsitz and forty other top mathematics students were published.

22 See chapter 2, note 1.

23 Ropolyi, "Lukács and Lakatos." Ropolyi also makes an educated guess that Lakatos attended Lukács's lectures on Hegel's *PhS* in 1945 or 1946.

24 Gábor Kutrovátz, "Imre Lakatos: Modern Physics, Modern Society" (forthcoming). Karácsony evaluated the dissertation on July 8, 1947, and wrote: "I got interested in the foregoing scientific activities of this young man, and not least because I read most of them at the moment they were published. Now I see all of Imre Lakatos' work in unity, and I deem that it comes up to the standard. His dissertation is not a sudden idea, it was matured by two previous publications, both in very serious journals. The first was published in *Athenaeum* [a journal of the Hungarian Academy of Sciences] under the title *A fizikai idealizmus bírálata* ['Criticism of Idealism in Physics'], and the second came out in a thick volume written to teachers: *Továbbképzés és demokrácia* [*Continuing Higher Education and Democracy*], entitled *Modern fizika, modern társadalo* ['Modern Physics, Modern Society']" (quoted in Kutrovátz). On the missing thesis, see Long, "Lakatos in Hungary," 271.

25 Quoted in Congdon, "Possessed," 281.

26 Personal communication, 1985.

27 For Lakatos's familiarity with *DR*, see Ropolyi, "Lukács and Lakatos."

28 From his essay "Recollections of Eötvös College's Last Years," quoted in Congdon, "Possessed," 285.

29 Keresztury and the yearbook, quoted in Congdon, "Possessed," 285, 287.

30 See Congdon, ibid.; and Long, "Lakatos in Hungary," 278.

31 See Jancis Long, "Lakatos in Hungary," 275.

32 Congdon, "Possessed," 289 n. 53. On the Recsk labor camp, see George Faludy, *My Happy Days in Hell,* trans. Kathleen Szász (Ontario: Totem Press, 1985), part 5.

33 See Alfred Rényi, *Dialogues on Mathematics* (San Francisco: Holden-Day, 1967).

34 "It is the basis of scientific education to train students and research students to respect facts, to demand exact thinking and proofs. Stalinism, on the other hand, branded these very demands as 'bourgeois objectivism.' Under the banner of party-

minded science, a large (even, we could say, world-scale) attempt has been made to create fact free and proof free scholarship or science. (For example, Lysenko's and Lepichinskaia's biology.) The extermination of facts was often carried out under the pretext of a 'Marxist' fight against empiricism—an invisible and frequently non-existent 'salient feature' was given first importance over the miserable and mostly unpalatable 'phenomena.' The victims of this fervid fight against formalism included logic, and many branches of applied mathematics (biometrics, econometrics) were anathematized. Dialectics was corrupted into scholastic sophism. The history of science indicates that we ought to teach the future scholar to be modest, to be humble in his scientific claims, to be averse to all kinds of fanaticism. He ought to learn that what he does not understand, or disapproves of, still has a right to exist, and that no scientific theory, no theorem can conclude anything finally, in the history of science. . . . New, hitherto unfamiliar chapters ought to be included in pedagogical textbooks, such as 'Methods for stimulating curiosity and developing it into interest,' 'How to teach to think scientifically,' 'How to teach people respect for facts' and—God forbid!—'How to teach people to doubt.' . . . [I]t would be good if our pedagogical textbooks devoted a chapter to '*How to teach respect for the right to dissent.*' " ("*Tudományra Nevelésröl*"/"On Rearing Scholars," quoted in Imre Lakatos and Paul Feyerabend, *For and Against Method,* ed. Matteo Motterlini [Chicago: University of Chicago Press, 1999], 375–81). As pointed out by Ropolyi and Kutrovátz, Lakatos's interest in scientific education and his views on freedom of thought appear consistently in his essays following the war.

35 Quoted in Congdon, "Possessed," 291.

36 Mária Zimán, *Betúsírko Évának: 1925–1944* (*A Memorial of Words for Eva: 1925–1944*) (privately published, 1989). I am grateful to Jancis Long for providing me with an English translation by Hajnal Csatorday.

37 See chapter 5.

38 When Nagy spoke to the crowds on October 24, again as prime minister, he began his speech with the greeting "Comrades," but was instantly booed. This account of Hungarian writers and intellectuals should not be attributed to the populace at large.

39 See G. W. F. Hegel, *Hegel's Science of Logic,* trans. A. V. Miller (New York: Humanities Press, 1976), on Spinoza: "The genuine refutation must penetrate the opponent's stronghold and meet him on his own ground; no advantage is gained by attacking him somewhere else and defeating him where he is not" (581). Then Lakatos: "While being addicted to his [Carnap's] own system of ideas, he never follows up the enemy in hostile terrain. . . . '*Live and let live*' is not a good rule for the dialectic of intellectual progress. If one does not follow up a critical clash to the bitter end, one may leave uncriticized not only the adversary but also oneself: for the best way to understand critically one's own position is through the relentless criticism of contrary positions" (*MSE* 136 n. 1).

40 George Orwell, *1984* (New York: Signet, 1983), 176. For a detailed study of Communist language using Orwell as a guide, see John Young, *Totalitarian Language* (Charlottesville: University Press of Virginia, 1991). Areas of interesting comparisons with Lakatos's writing style include "Ism in Communist Propaganda"

and "Ironic Quotation Marks" (173 ff.). Young reports that a Hungarian translation of *1984* was secretly circulated among members of the Petőfi Circle in 1956.

41 Orwell, *1984*, 176.

42 Ibid., 177.

43 Ibid., 175–76.

44 On disguised opposition, see Váli, *Rift and Revolt*, 202 ff.; and Young, *Totalitarian Language*, 182–86. For an example from a speech by Nagy, see Imre Kovács, ed., *Facts about Hungary* (New York: Hungarian Committee, 1959), 72–75.

45 Georg Lukács, *The Meaning of Contemporary Realism*, trans. John and Necke Mander (London: Merlin, 1963), 9.

46 Orwell, *1984*, 177.

47 Ibid., 35.

48 See Hans-Georg Gadamer, "Hegel's Inverted World," in *Hegel's' Dialectic: Five Hermeneutical Studies*, trans. P. Christopher Smith (New Haven, Conn.: Yale University Press, 1976). The inverted world is developed in the section "Force and the Understanding" in *PhS* (79–103).

49 "It is an enchanted, perverted, topsy-turvy world, in which Monsieur le Capital and Madame la Terre do their ghost-walking as social characters and at the same time directly as mere things" (Marx, *Capital*, 3: 830).

50 See Tibor Fischer's novel, *Under the Frog* (New York: New Press, 1992): Even "the Writers' Union, the home of moral malnutrition, was at it, suddenly disclaiming all the things they had written in the last few years. The Union had pulled its head out of Rákosi's arse and now stood blinking in the daylight" (197). See also 235, on a book called *Testimony* on U.S. atrocities in Korea by Méray: "The only things in the book that weren't downright lies were the author's name and the commas."

51 Georg Lukács's view, as he said, in *Record of a Life*, ed. István Eörsi, trans. Rodney Livingstone (London: Verso, 1983), is that "Marx established that historicity is the fundamental concept of social being, and as such of all beings. This I hold to be the most important part of Marxian theory. . . . Traditional philosophy conceived of a system of categories which included the categories of history along with others. In the Marxist system of categories every object is furnished from the outset with attributes, with thinghood and with a categorical existence. . . . It is not the case that history unfolds within the system of categories, but rather that history is the system of categories in the process of change" (142). Thus, for Lukács, the rejection of history as a primary philosophical category is equivalent to the rejection of Marx and reason.

52 Theodor Adorno, "Reconciliation under Duress," in *Aesthetics and Politics*, ed. Ronald Taylor (London: New Left Books, 1977), 152; George Lichtheim, *Georg Lukács* (New York: Viking, 1970), 115; and George Steiner, *Language and Silence* (New York: Athenaeum, 1970), 329. Lukács's student Ágnes Heller describes *DR* as a "two-volume pamphlet" and "demonology," but that "Lukács was not alone in his attempt to attribute all the horrors of the twentieth century to various philosophies. The belief that there are no innocent ideas was very widespread. Popper in his *Open Society*, or [Max] Horkheimer and Adorno in their *Dialectic of Enlightenment*, trod the same road of *a posteriori* construction. . . . Horkheimer and Adorno (in contrast to Lukács) saw the root of the evil in a certain interpreta-

tion of rationalism. All of these works were characterized by a kind of demonology, for they split philosophies into good and bad ones. It is sufficient to glance at Popper's title: 'open society' embodies the Good and the opposite tendency is designated as the 'enemy' in order to show these demonological tendencies" ("Lukács' Later Philosophy," in *Lukács Reappraised,* ed. Agnes Heller [New York: Columbia University Press, 1983], 179). Ferenc Fehér also compares the trajectories of DR, Popper, and Lukács, and how in the *Dialectic of Enlightenment* "the famous Lukácsian *Leitmotiv* of the 1920s, the exposure of rationalism as 'the spirit of capitalism' . . . had been transformed into a real philosophical perversion" ("Lukács in Weimar," in *Lukács Reappraised,* ed. Ágnes Heller [New York: Columbia University Press, 1983], 88). Popper's *Open Society,* therefore, is as important for appreciating Lakatos's work and its Hegelian-Marxist vision as is LSCD.

53 Leszek Kolakowski, *Main Currents of Marxism,* trans. P. S. Falla (New York: Oxford University Press, 1981), 3: 285. The chapter is titled "György Lukács: Reason in the Service of Dogma."

54 Paul Feyerabend, "Imre Lakatos," *British Journal for the Philosophy of Science* 26 (1975): 3.

55 Lukács, *Record of a Life,* 101–5.

56 Georg Lukács, "On the Responsibility of Intellectuals," in *Marxism and Human Liberation: Essays on History, Culture, and Revolution by Georg Lukács,* ed. E. San Juan (New York: Delta, 1973), 268; for a similar passage, see also DR, 90.

57 Kolakowski, *Main Currents of Marxism,* 3: 253. On Lakatos as a "Trojan horse," see also introduction, note 17. On support for Lukács as "an irregular and irreverent 'partisan,'" see Fehér, "Lukács in Weimar," 77.

58 See Lukács's frightening remarks on Party discipline and renouncing personal freedoms in the final chapter of H&CC, "Towards a Methodology of the Problem of Organization" (for example, 315 ff.).

59 Lukács, *Record of a Life,* 65.

60 Quoted in Lasky, *The Hungarian Revolution,* 159. Note also Kolakowski's characterization of Lukács-Naphta: "A highly intelligent character who needs authority, finds it, and renounces his own personality for its sake. Lukács in fact was a true intellectual, a man of immense culture . . . but one who craved intellectual security and could not endure the uncertainties of a sceptical or empirical outlook. In the Communist Party he found what many intellectuals need: absolute certainty in defiance of facts" (*Main Currents of Marxism,* 3: 306). Kolakowski complains that in the "late 1950s," Lukács was "one of the most timid and cautious critics of Stalinism" (ibid.), but I am unsure what should be expected of one of the last Hungarians to see Nagy alive, in late 1950s' Hungary no less.

61 For Stephen Toulmin's reasonably puzzled comment, see his "History, Praxis, and the 'Third World': Ambiguities in Lakatos' Theory of Methodology," in *Essays in Memory of Imre Lakatos,* ed. Robert Cohen et al. (Boston: D. Reidel, 1976).

62 The letter is reprinted as chapter 12 in MSE.

Bibliography

Aczél, Tamás. "The Mythology of True Consciousness." In *Ten Years After: A Commemoration of the Tenth Anniversary of the Hungarian Revolution.* London: Macgibbon and Kee, 1966.

Aczél, Tamás, and Tibor Méray. *The Revolt of the Mind: A Case Study of Intellectual Resistance behind the Iron Curtain.* New York: Praeger, 1958.

Adorno, Theodor. "Reconciliation under Duress." In *Aesthetics and Politics,* edited by Ronald Taylor. London: New Left Books, 1977.

Arato, Andrew, and Paul Breines. *The Young Lukács and the Origins of Western Marxism.* New York: Continuum/Seabury, 1979.

Aristotle. *The Complete Works of Aristotle.* Edited by Jonathan Barnes. 2 vols. New York: Oxford University Press, 1984.

Avineri, Shlomo. *The Social and Political Thought of Karl Marx.* Cambridge, U.K.: Cambridge University Press, 1971.

Backhouse, Roger. *Explorations in Economic Methodology: From Lakatos to Empirical Philosophy of Science.* New York: Routledge, 1998.

——, ed. *New Directions in Economic Methodology.* New York: Routledge, 1994.

Bartley, William, III. "On Imre Lakatos." In *Essays in Memory of Imre Lakatos,* edited by Robert Cohen et al. Boston: D. Reidel, 1976.

Bergsten, Gunilla. *Thomas Mann's "Doctor Faustus": The Sources and Structure of the Novel.* Translated by Krishna Winston. Chicago: University of Chicago Press, 1969.

Birkhoff, Garrett, ed. *A Source Book in Classical Analysis.* Cambridge, Mass.: Harvard University Press, 1973.

Blaug, Mark. "Kuhn versus Lakatos *or* Paradigms versus Research Programs in the History of Economics." In *Method and Appraisal in Economics,* edited by Spiro Latsis. New York: Cambridge University Press, 1976.

——. *Economic Theory in Retrospect.* 3d ed. New York: Cambridge University Press, 1978.

——. *Methodological Appraisal of Marxian Economics.* Amsterdam: North-Holland, 1980.

Bochner, Salomon. *The Role of Mathematics in the Rise of Science.* Princeton, N.J.: Princeton University Press, 1966.

Boolos, George. *The Unprovability of Consistency.* New York: Cambridge University Press, 1979.

Boyer, Carl. *The History of the Calculus and Its Conceptual Development.* 1949. Reprint, New York: Dover, 1959.
Brush, Stephen. *The Kind of Motion We Call Heat: A History of the Kinetic Theory of Gases in the Nineteenth Century.* 2 vols. Amsterdam: North-Holland, 1976.
Bubner, Rüdiger. *Modern German Philosophy.* Translated by Eric Matthews. New York: Cambridge University Press, 1981.
Buchwald, Jed. *The Rise of the Wave Theory of Light: Optical Theory and Experiment in the Early Nineteenth Century.* Chicago: University of Chicago Press, 1989.
——, ed. *Scientific Practice: Theories and Stories of Doing Physics.* Chicago: University of Chicago Press, 1995.
Burnyeat, Myles. "Protagoras and Self-Refutation in Later Greek Philosophy." *Philosophical Review* 85 (1976): 44–69.
——. "Can the Skeptic Live His Skepticism?" In *The Skeptical Tradition,* edited by Myles Burnyeat. Berkeley: University of California Press, 1983.
——. "The Sceptic in His Place and Time." In *Philosophy in History: Essays on the Historiography of Philosophy,* edited by Richard Rorty et al. New York: Cambridge University Press, 1984.
——, ed. *The Skeptical Tradition.* Berkeley: University of California Press, 1983.
Cartwright, Nancy. *How the Laws of Science Lie.* New York: Oxford University Press, 1983.
Clark, Peter. "Atomism versus Thermodynamics." In *Method and Appraisal in the Physical Sciences: The Critical Background to Science, 1850–1900,* edited by Colin Howson. New York: Cambridge University Press, 1976.
Close, Frank, et al. *The Particle Explosion.* New York: Oxford University Press, 1987.
Cohen, Paul. *Set Theory and the Continuum Hypothesis.* Reading, Pa.: W. A. Benjamin, 1966.
Cohen, Robert, et al., eds. *Essays in Memory of Imre Lakatos.* Boston: D. Reidel, 1976.
Colletti, Lucio. *Marxism and Hegel.* Translated by Lawrence Garner. London: Verso, 1973.
——. "Marxism and the Dialectic." *New Left Review* 93 (1975): 3–29.
——. "A Political and Philosophical Interview." In *Western Marxism: A Critical Reader,* edited by New Left Review. London: Verso, 1978.
Congdon, Lee. "Possessed: Imre Lakatos' Road to 1956." *Contemporary European History* 6 (1997): 279–94.
Couissin, Pierre, "The Stoicism of the New Academy." In *The Skeptical Tradition,* edited by Myles Burnyeat. Berkeley: University of California Press, 1983.
Crampton, R. J. *Eastern Europe in the Twentieth Century.* New York: Routledge, 1994.
Dauben, Joseph. "The Trigonometric Background to Georg Cantor's Theory of Sets." *Archive for History of Exact Sciences* 7 (1971): 181–216.
——. *Georg Cantor: His Mathematics and Philosophy of the Infinite.* Princeton, N.J.: Princeton University Press, 1979.
Davis, Phillip, and Reuben Hersh. *The Mathematical Experience.* Boston: Houghton Mifflin, 1981.
Dawson, John. "The Reception of Gödel's Incompleteness Theorems." In *Proceedings of the 1984 Biennial Meeting of the Philosophy of Science Association: Volume 2,* edited by Peter Asquith and Philip Kitcher. East Lansing, Mich.: Philosophy of Science Association, 1985.

di Giovanni, George, and H. S. Harris, eds. *Between Kant and Hegel: Texts in the Development of Post-Kantian Idealism.* Albany: State University of New York Press, 1985.

Dobb, Maurice. *Theories of Value and Distribution Since Adam Smith.* New York: Cambridge University Press, 1973.

Dove, Kenley. "Hegel's Phenomenological Method." In *New Studies in Hegel's Philosophy,* edited by Warren Steinkraus. New York: Holt, Rinehart & Winston, 1971.

Duhem, Pierre. *The Aim and Structure of Physical Theory.* Translated by Philip Weiner. 1914. Reprint, Princeton, N.J.: Princeton University Press, 1954.

Eatwell, John, et al., eds. *The New Palgrave Marxian Economics.* New York: W. W. Norton, 1987.

Einstein, Albert. "On the Electrodynamics of Moving Bodies." In *The Principle of Relativity,* edited by Hendrik Lorentz et al. Translated by W. Perrett and G. B. Jeffery. New York: Dover, 1952.

Eiseley, Loren. *Darwin's Century.* Garden City, N.Y.: Anchor Books, 1962.

Elkana, Yehuda, ed. *The Interaction between Science and Philosophy.* Atlantic Highlands, N.J.: Humanities Press, 1974.

Faludy, George. *My Happy Days in Hell.* Translated by Kathleen Szász. Ontario: Totem Press, 1985.

Feferman, Solomon. "Arithmetization of Metamathematics in a General Setting." *Fundamenta Mathematicae* 49 (1960): 35–92.

——. "My Route to Arithmetization," *Theoria* 63 (1997): 168–81.

——. *In the Light of Logic.* New York: Oxford University Press, 1999.

Fehér, Ferenc. "Lukács in Weimar." In *Lukács Reappraised,* edited by Ágnes Heller. New York: Columbia University Press, 1983.

Feyerabend, Paul. "Against Method." In *Minnesota Studies in the Philosophy of Science IV,* edited by M. Radner and S. Winokur. Minneapolis: University of Minnesota Press, 1970.

——. *Against Method.* London: New Left Books, 1975.

——. "Imre Lakatos." *British Journal for the Philosophy of Science* 26 (1975): 1–18.

——. *Science in a Free Society.* London: New Left Books, 1978.

——. *Problems of Empiricism: Philosophical Papers, Volume 2.* New York: Cambridge University Press, 1981.

——. *Killing Time: The Autobiography of Paul Feyerabend.* Chicago: University of Chicago Press, 1995.

Fischer, Tibor. *Under the Frog.* New York: New Press, 1992.

Fogelin, Robert. "The Tendency of Hume's Skepticism." In *The Skeptical Tradition,* edited by Myles Burnyeat. Berkeley: University of California Press, 1983.

Forster, Michael. *Hegel and Skepticism.* Cambridge, Mass.: Harvard University Press, 1989.

——. *Hegel's Idea of a Phenomenology of Spirit.* Chicago: University of Chicago Press, 1998.

Fried, Michael. "The Skeptic's Two Kinds of Assent and the Question of the Possibility of Knowledge." In *Philosophy in History: Essays on the Historiography of Philosophy,* edited by Richard Rorty et al. New York: Cambridge University Press, 1984.

Funtowicz, Silvio, and J. R. Ravetz. *Uncertainty and Quality in Science for Policy.* Dordrecht, Netherlands: Kluwer, 1990.

Gadamer, Hans-Georg. *Truth and Method.* Translated by Joel Weinsheimer and Donald Marshall. New York: Seabury, 1975.

——. *Hegel's Dialectic: Five Hermeneutical Studies.* Translated by P. Christopher Smith. New Haven, Conn.: Yale University Press, 1976.

Galison, Peter. "Context and Constraints." In *Scientific Practice: Theories and Stories of Doing Physics,* edited by Jed Buchwald. Chicago: University of Chicago Press, 1995.

——. *Image and Logic: A Materialist Culture of Microphysics.* Chicago: University of Chicago Press, 1997.

Gati, Charles. *Hungary and the Soviet Bloc.* Durham, N.C.: Duke University Press, 1986.

Giard, Luce. Introduction to *Histoire et Methodologie des Sciences: Programmes de recherche et reconstruction rationnelle.* Paris: Presses Universitaires de France, 1994.

Gillman, Joseph. *The Falling Rate of Profit.* London: Denis Dobson, 1957.

Gödel, Kurt. "What Is Cantor's Continuum Problem?" In *The Philosophy of Mathematics,* edited by Paul Benacerraf and Hilary Putnam. 1947. Reprint, Englewood Cliffs, N.J.: Prentice-Hall, 1964.

——. *Kurt Gödel: Collected Works, Volume I.* Edited by Solomon Feferman et al. New York: Oxford University Press, 1986.

Goethe, Johann Wolfgang von. *Goethe's Faust.* Translated by Walter Kaufmann. Garden City, N.Y.: Doubleday, 1961.

——. *Goethe's Conversations with Eckermann, 1823–1832.* Translated by John Oxenford. San Francisco: North Point Press, 1984.

——. *Wilhelm Meister* 1778–1796. Translated by Thomas Carlyle. 3 vols. New York: E. P. Dutton, 1824.

Goldberg, Stanley. *Understanding Relativity: Origin and Impact of a Scientific Revolution.* Boston: Birkhauser, 1984.

Goldfarb, Warren. "Logic in the Twenties: The Nature of the Quantifier." *Journal of Symbolic Logic* 44 (1979): 351–68.

Gould, Stephen J. *Ontogeny and Phylogeny.* Cambridge, Mass.: Harvard University Press, 1977.

——. *The Mismeasure of Man.* New York: W. W. Norton, 1981.

Grattan-Guinness, Ivor. *The Development of Mathematical Analysis from Euler to Riemann.* Cambridge, Mass.: MIT Press, 1970.

——. "The Emergence of Mathematical Analysis and Its Foundational Progress, 1780–1880." In *From the Calculus to Set Theory: 1630–1910,* edited by Ivor Grattan-Guinness. London: Duckworth, 1980.

Gray, Ronald. *Goethe: A Critical Introduction.* New York: Cambridge University Press, 1967.

Habermas, Jürgen. *Knowledge and Human Interests.* Translated by Jeremy Shapiro. Boston: Beacon Press, 1971.

Hacking, Ian. "Imre Lakatos' Philosophy of Science." *British Journal for the Philosophy of Science* 30 (1979): 381–410.

——. *Representing and Intervening.* New York: Cambridge University Press, 1983.

Hallet, Michael. "Towards a Theory of Mathematical Research Programmes." *British Journal for the Philosophy of Science* 30 (1979): 1–25, 135–59.

Hanson, Norwood. *Patterns of Discovery.* New York: Cambridge University Press, 1969.

Harris, H. S. *Hegel's Development: Towards the Sunlight, 1770–1801.* New York: Oxford University Press, 1972.

Hawkins, Thomas. *Lebesgue's Theory of Integration: Its Origins and Development.* Madison: University of Wisconsin Press, 1970.

Heath, Thomas. *A History of Greek Mathematics.* 2 vols. 1921. Reprint, New York: Dover, 1981.

Hegel, G. W. F. Preface to *Phenomenology of Spirit.* In *Hegel: Texts and Commentary,* trans. Walter Kaufmann. Garden City, N.Y.: Anchor-Doubleday, 1965.

——. Introduction to *History of Philosophy.* In *Hegel's Idea of Philosophy,* edited and translated by Quentin Lauer. New York: Fordham University Press, 1974.

——. *Lectures on the History of Philosophy.* Translated by E. S. Haldane and F. H. Simson. 3 vols. 1892. Reprint, New York: Humanities Press, 1974.

——. *Hegel's Logic.* Pt. 1 of *Hegel's Encyclopedia,* translated by William Wallace. 1812–1816. Reprint, Oxford: Oxford University Press, 1976.

——. *Lectures on the Philosophy of World History, Introduction: Reason in History.* Translated by H. B. Nisbet. New York: Cambridge University Press, 1975.

——. *Hegel's Science of Logic.* Translated by A. V. Miller. 1812–1816. New York: Humanities Press, 1976.

——. *Hegel's Phenomenology of Spirit.* Translated by A. V. Miller. 1807. Reprint, Oxford: Oxford University Press, 1979.

——. *Hegel's Philosophy of Right.* Translated by T. M. Knox. 1821. Reprint, New York: Oxford University Press, 1979.

Heller, Agnes, "Lukács' Later Philosophy." In *Lukács Reappraised,* edited by Ágnes Heller. New York: Columbia University Press, 1983.

Hintikka, Jaakko, and Unto Remes. *The Method of Analysis: Its Geometrical Origin and Its General Significance.* Boston: D. Reidel, 1974.

Hirosige, Tetu. "The Ether Problem, the Mechanistic Worldview, and the Origins of the Theory of Relativity." *Historical Studies in the Physical Sciences* 7 (1976): 3–81.

Hodos, George. *Show Trials: Stalinist Purges in Eastern Europe, 1948–1954.* New York: Praeger, 1984.

Holton, Gerald. *Thematic Origins of Scientific Thought: Kepler to Einstein.* Rev. ed. Cambridge, Mass.: Harvard University Press, 1973.

Howson, Colin, ed. *Method and Appraisal in the Physical Sciences: The Critical Background to Science, 1850–1900.* New York: Cambridge University Press, 1976.

Hume, David. *Dialogues concerning Natural Religion.* Edited by Norman Kemp Smith. 1751–1755. Reprint, Indianapolis: Bobbs-Merrill, 1947.

——. *An Enquiry concerning Human Understanding.* Edited by L. A. Selby-Bigge and P. H. Nidditch. 3d ed. 1777. Reprint, Oxford: Oxford University Press, 1975.

——. *A Treatise of Human Nature.* 1739. Reprint, Buffalo, N.Y.: Prometheus, 1992.

Hunt, E. K., and Mark Glick. "Transformation Problem." In *The New Palgrave Marxian Economics,* edited by John Eatwell et al. New York: W. W. Norton, 1987.

Ignotus, Paul. *Hungary.* New York: Praeger, 1972.

Ihde, Aaron. *The Development of Modern Chemistry.* New York: Harper and Row, 1965.

Jammer, Max. *The Conceptual Development of Quantum Mechanics.* New York: McGraw-Hill, 1966.

Jungnickel, Christa, and Russell McCormmach. *The Intellectual Mastery of Nature: Theoretical Physics from Ohm to Einstein.* 2 vols. Chicago: University of Chicago Press, 1986.

Kadvany, John. "A Mathematical Bildungsroman." *History and Theory* 28 (1989): 25–42.

——. "Dialectic and Diagonalization." *Inquiry* 34 (1991): 3–25.

——. "Taming Chance: Risk and the Quantification of Uncertainty." *Policy Sciences* 29 (1996): 1–27.

Kant, Immanuel. *Critique of Pure Reason.* Translated by Norman Kemp Smith. 1781–1787. Reprint, New York: St. Martin's Press, 1965.

Kaufmann, Walter. *From Shakespeare to Existentialism.* Garden City, N.Y.: Anchor-Doubleday, 1959.

——. *Hegel: A Reinterpretation.* Garden City, N.Y.: Anchor-Doubleday, 1965.

King, David. *The Commissar Vanishes: The Falsification of Photographs and Art in Stalin's Russia.* New York: Henry Holt, 1997.

Király, Béla, et al., eds. *The First War between Socialist States: The Hungarian Revolution of 1956 and Its Impact.* New York: Brooklyn College Press, 1984.

Kneebone, G. T. *Mathematical Logic and the Foundations of Mathematics.* London: Van Nostrand, 1963.

Koetsier, Teun. *Lakatos' Philosophy of Mathematics: An Historical Approach.* Amsterdam: North-Holland, 1991.

Kolakowski, Leszek. *Main Currents of Marxism.* Translated by P. S. Falla. 3 vols. New York: Oxford University Press, 1981.

Konrád, Georg. *The Loser.* Translated by Ivan Sanders. New York: Harcourt Brace Jovanovich, 1982.

Kovács, Imre, ed. *Facts about Hungary.* New York: Hungarian Committee, 1959.

Kreisel, Georg. "On a Problem of Henkin's." *Indagationes Mathematicae* 15 (1953): 405–6.

——. "Mathematical Significance of Consistency Proofs." *Journal of Symbolic Logic* 23 (1958): 155–82.

——. "A Survey of Proof Theory." *Journal of Symbolic Logic* 33 (1968): 321–88.

Krell, David. "The Mole: Philosophic Burrowings in Kant, Hegel, and Nietzsche." In *Why Nietzsche Now?* edited by Daniel O'Hara. Bloomington: Indiana University Press, 1985.

Kuhn, Thomas. *The Copernican Revolution: Planetary Astronomy in the Development of Western Thought.* Cambridge, Mass.: Harvard University Press, 1957.

——. "Logic of Discovery or Psychology of Research?" In *Criticism and the Growth of Knowledge,* edited by Imre Lakatos and Alan Musgrave. New York: Cambridge University Press, 1970.

——. "Reflections on My Critics." In *Criticism and the Growth of Knowledge,* edited by Imre Lakatos and Alan Musgrave. New York: Cambridge University Press, 1970.

——. *The Structure of Scientific Revolutions.* 2d ed. Chicago: University of Chicago Press, 1970.

——. *The Essential Tension: Selected Studies in Scientific Tradition and Change.* Chicago: University of Chicago Press, 1977.

Kutrovátz, Gábor. "Imre Lakatos: Modern Physics, Modern Society." Forthcoming. Translation of "*Modern fizik, modern társadalom,*" by Imre Lakatos. In *Továbbképzés és demokrácia,* edited by Gábor Kimény. Budapest: Vallás és Közoktatásügyi Minisztérium nevelésügyi osztálya, 1947.

Lakatos, Imre. *Proofs and Refutations: The Logic of Mathematical Discovery.* Edited by John Worrall and Elie Zahar. New York: Cambridge University Press, 1976.

——. *Mathematics, Science, and Epistemology: Philosophical Papers, Volume II.* Edited by John Worrall and Gregory Currie. New York: Cambridge University Press, 1978.

——. *The Methodology of Scientific Research Programmes: Philosophical Papers, Volume I.* Edited by John Worrall and Gregory Currie. New York: Cambridge University Press, 1978.

Lakatos, Imre, and Paul Feyerabend. *For and Against Method.* Edited by Matteo Motterlini. Chicago: University of Chicago Press, 1999.

Lakatos, Imre, and Alan Musgrave, eds. *Criticism and the Growth of Knowledge.* New York: Cambridge University Press, 1970.

Lakatos, Imre, and Elie Zahar. "Why Did Copernicus' Research Programme Supersede Ptolemy's?" In *The Methodology of Scientific Research Programmes: Philosophical Papers I,* edited by John Worrall and Gregory Currie. New York: Cambridge University Press, 1978.

Lasky, Melvin, ed. *The Hungarian Revolution: A White Book.* New York: Praeger, 1957.

Latsis, Spiro, ed. *Method and Appraisal in Economics.* New York: Cambridge University Press, 1976.

Lichtheim, George. *Marxism.* 2d ed. New York: Praeger, 1965.

——. *The Origins of Socialism.* New York: Praeger, 1969.

——. *Georg Lukács.* New York: Viking, 1970.

Litván, György, ed. *The Hungarian Revolution of 1956: Reform, Revolt, and Repression, 1953–1963.* Translated by János Bak and Lyman Legters. New York: Longman, 1996.

Long, Jancis. "Lakatos in Hungary." *Philosophy of the Social Sciences* 28 (1998): 245–311.

Lorentz, Hendrik, et al. *The Principle of Relativity,* 1895–1918. Translated by W. Perrett and G. B. Jeffery. New York: Dover, 1952.

Lukács, Georg. *The Meaning of Contemporary Realism.* 1957. Translated by John and Necke Mander. Reprint, London: Merlin, 1963.

——. *Goethe and His Age.* Translated by Robert Anchor. 1947. Reprint, London: Merlin Press, 1968.

——. *History and Class Consciousness: Studies in Marxist Dialectics.* Translated by Rodney Livingstone. 1923. Reprint, Cambridge, Mass.: MIT Press, 1971.

——. "On the Responsibility of Intellectuals." In *Marxism and Human Liberation: Essays on History, Culture, and Revolution by Georg Lukács,* edited by E. San Juan, 1955. Reprint, New York: Delta, 1973.

——. *The Young Hegel: Studies in the Relations between Dialectics and Economics.* Translated by Rodney Livingstone. 1948. Reprint, Cambridge, Mass.: MIT Press, 1976.

——. *The Destruction of Reason.* Translated by Peter Palmer. 1954. Reprint, Atlantic Highlands, N.J.: Humanities Press, 1981.

——. *Record of a Life.* Edited by István Eörsi. Translated by Rodney Livingstone. London: Verso, 1983.

Maddy, Penelope. "Believing the Axioms, I, II." *Journal of Symbolic Logic* 53 (1988): 481–511, 736–64.

Mann, Thomas. *The Magic Mountain.* Translated by H. T. Lowe-Porter. New York: Vintage, 1969.

Marx, Karl. *Karl Marx: Early Writings.* Edited and translated by Tom Bottomore. New York: McGraw-Hill, 1963

——. *Theories of Surplus-Value.* Translated by Emil Burns. 3 vols. 1905–1910. Reprint, Moscow: Progress Publishers, 1963.

——. *Capital.* Edited by Friedrich Engels. Translated by Samuel Moore and Edward Aveling. 3d ed. 3 vols. 1867–1894. Reprint, New York: International Publishers, 1967.

——. *Grundrisse: Foundations of the Critique of Political Economy.* Translated by Martin Nicolaus. 1939. Reprint, New York: Vintage, 1973.

——. "The Eighteenth Brumaire of Louis Bonaparte." In *The Marx-Engels Reader,* edited by Robert Tucker. 2d ed. 1852. Reprint, New York: W. W. Norton, 1978.

McCormmach, Russell. "Einstein, Lorentz, and the Electron Theory." *Historical Studies in the Physical Sciences* 2 (1970): 41–87.

——. "H. A. Lorentz and the Electromagnetic View of Nature." *Isis* 61 (1970): 459–97.

Meek, Ronald. *Studies in the Labour Theory of Value.* 2d ed. London: Lawrence and Wishart, 1973.

Miller, Arthur. *Albert Einstein's Special Theory of Relativity: Emergence (1905) and Early Interpretation (1905–1911).* Reading, Mass.: Addison-Wesley, 1981.

Monna, A. F. "The Concept of Functions in the Nineteenth and Twentieth Centuries, in Particular with Regard to the Discussions between Baire, Borel, and Lebesgue." *Archive for History of Exact Sciences* 9 (1972): 57–84.

Moore, Gregory. *Zermelo's Axiom of Choice: Its Origins, Development, and Influence.* New York: Springer, 1982.

——. "The Emergence of a First-Order Logic." In *History and Philosophy of Modern Mathematics,* edited by William Aspray and Philip Kitcher. Minneapolis: University of Minnesota Press, 1988.

Morishima, Michio. "Marx in the Light of Modern Economic Theory." *Econometrica* 42 (1974): 611–32.

Morrison, Margaret. "Scientific Conclusions and Philosophical Arguments: An Inessential Tension." In *Scientific Practice: Theories and Stories of Doing Physics,* edited by Jed Buchwald. Chicago: University of Chicago Press, 1995.

Musgrave, Alan. "Method or Madness: Can the Methodology of Scientific Research Programmes Be Rescued from Epistemological Anarchism?" In *Essays in Memory of Imre Lakatos,* edited by Robert Cohen et al. Boston: D. Reidel, 1976.

——. "Why Did Oxygen Supplant Phlogiston? Research Programmes in the Chemical Revolution." In *Method and Appraisal in the Physical Sciences: The Critical Background to Science, 1850–1900,* edited by Colin Howson. New York: Cambridge University Press, 1976.

Orwell, George. *1984.* 1949. Reprint, New York: Signet, 1983.

Pais, Abraham. *"Subtle Is the Lord . . .": The Science and Life of Albert Einstein.* New York: Oxford University Press, 1982.

Paris, Jeff, and Leo Harrington. "A Mathematical Incompleteness in Peano Arithmetic." In *Handbook of Mathematical Logic,* edited by Jon Barwise. Amsterdam: North-Holland, 1977.

Pickering, Andrew. *The Mangle of Practice.* Chicago: University of Chicago Press, 1995.

Plato. *The Collected Dialogues.* Edited by Edith Hamilton and Huntington Cairns. Princeton, N.J.: Princeton University Press, 1961.

——. *The Theaetetus of Plato.* Translated by M. J. Levett. Edited by Myles Burnyeat. Indianapolis, Ind.: Hackett, 1990.

Popkin, Richard. *The History of Scepticism: From Erasmus to Spinoza.* 3d ed. Berkeley: University of California Press, 1979.

——. "David Hume and the Pyrrhonian Controversy." In *The High Road to Pyrrhonism,* edited by R. A. Watson and J. E. Force. San Diego, Calif.: Austin Hill Press, 1980.

——. "David Hume: His Pyrrhonism and His Critique of Pyrrhonism." In *The High Road to Pyrrhonism,* edited by R. A. Watson and J. E. Force. San Diego, Calif.: Austin Hill Press, 1980.

Popper, Karl. *The Open Society and Its Enemies.* 4th ed. 2 vols. 1943. Reprint, Princeton, N.J.: Princeton University Press, 1963.

——. *Conjectures and Refutations: The Growth of Scientific Knowledge.* New York: Harper Torchbooks, 1968.

——. *The Logic of Scientific Discovery.* 1959. Reprint, New York: Harper, 1968.

——. "Normal Science and Its Dangers." In *Criticism and the Growth of Knowledge,* edited by Imre Lakatos and Alan Musgrave. New York: Cambridge University Press, 1970.

——. *Objective Knowledge: An Evolutionary Approach.* Oxford: Oxford University Press, 1972.

——. "Lakatos on the Equal Status of Newton's and Freud's Theories." In *The Philosophy of Karl Popper,* edited by Paul Schlipp. Vol. 2 LaSalle, Ill.: Open Court, 1974.

Postone, Moishe. *Time, Labor, and Social Domination: A Reinterpretation of Marx's Critical Theory.* New York: Cambridge University Press, 1996.

Preston, John. *Feyerabend: Philosophy, Science, and Society.* Cambridge, U.K.: Polity Press, 1997.

Purrington, Robert. *Physics in the Nineteenth Century.* New Brunswick, N.J.: Rutgers University Press, 1997.

Putnam, Hilary. "The Logic of Quantum Mechanics." In *Mathematics, Matter, and Method: Philosophical Papers, Volume I.* New York: Cambridge University Press, 1975.

Quine, Willard Van Orman. "Two Dogmas of Empiricism." In *From a Logical Point of View.* Cambridge, Mass.: Harvard University Press, 1961.

Rabinow, Paul. *Making PCR: A Story of Biotechnology.* Chicago: University of Chicago Press, 1996.

Ravetz, J. R. "Vibrating Strings and Arbitrary Functions." In *The Logic of Personal Knowledge: Essays Presented to Michael Polányi on His Seventieth Birthday,* edited by M. Grene. Glencoe, Ill.: Free Press, 1961.

Reiss, Hans. *Goethe's Novels.* New York: St. Martin's Press, 1969.
Rényi, Alfréd. *Dialogues on Mathematics.* San Francisco: Holden-Day, 1967.
Ricardo, David. *Principles of Political Economy and Taxation.* 1817. Reprint, Buffalo, N.Y.: Prometheus, 1996.
Roemer, John. "Marxian Value Analysis." In *The New Palgrave Marxian Economics,* edited by John Eatwell et al. New York: W. W. Norton, 1987.
Ropolyi, László. "Lukács and Lakatos." In *The Philosophy of Imre Lakatos: Its Roots, Content, and Limitations,* edited by Michael Stöltzner et al. Dordrecht, Netherlands: Kluwer, forthcoming.
Rosdolsky, Roman. *The Making of Marx's "Capital."* Translated by Pete Burgess. London: Pluto Press, 1977.
Sarkar, Husain. *A Theory of Method.* Berkeley: University of California Press, 1983.
Schmidt, Alfred. *The Concept of Nature in Marx.* Translated by Ben Fowkes. London: New Left Books, 1971.
Sedly, David. "The Motivation of Greek Skepticism." In *The Skeptical Tradition,* edited by Myles Burnyeat. Berkeley: University of California Press, 1983.
Segrè, Emilio. *From Falling Bodies to Radio Waves.* New York: W. H. Freeman, 1984.
Sextus Empiricus. *Outlines of Pyrrhonism.* Translated by R. C. Bury. 4 vols. Cambridge, U.K.: Loeb Classical Library, 1939.
Shapin, Steven. *A Social History of Truth: Civility and Science in Seventeenth-Century England.* Chicago: University of Chicago Press, 1994.
Simpson, Stephen. *Subsystems of Second Order Arithmetic.* New York: Springer, 1999.
Smith, Adam. *An Inquiry into the Nature and Causes of the Wealth of Nations.* 1776. Reprint, New York: Modern Library, 1937.
Smith, John. "Hegel's Critique of Kant." *Review of Metaphysics* 26 (1973): 438–60.
Soros, George. *The Alchemy of Finance.* New York: Simon and Schuster, 1987.
Steedman, Ian. *Marx after Sraffa.* London: Verso, 1977.
Steiner, George. *Language and Silence.* New York: Athenaeum, 1970.
——. *Extraterritorial: Papers on Literature and the Language Revolution.* New York: Atheneum, 1976.
Stough, Charlotte. *Greek Skepticism.* Berkeley: University of California Press, 1969.
Striker, Gisela. "The Ten Tropes of Aenesidemus." In *The Skeptical Tradition,* edited by Myles Burnyeat. Berkeley: University of California Press, 1983.
Swenson, Loyd. *The Ethereal Aether: A History of the Michelson-Morley-Miller Aether-Drift Experiments, 1880–1930.* Austin: University of Texas Press, 1972.
Szabó, Árpád. *The Beginnings of Greek Mathematics.* Translated by A. M. Ungar. Dordrecht, Netherlands: D. Reidel, 1978.
Szász, Béla. *Volunteers for the Gallows: Anatomy of a Show-Trial.* Translated by Kathleen Szász. New York: W. W. Norton, 1971.
Szendry, Thomas. "Philosophy in Hungary before and after 1956: From Staunch Orthodoxy to Limited Revisionism." In *The First War between Socialist States: The Hungarian Revolution of 1956 and Its Impact,* edited by Béla Király et al. New York: Brooklyn College Press, 1984.
Toulmin, Stephen. "History, Praxis, and the 'Third World': Ambiguities in Lakatos' Theory of Methodology." In *Essays in Memory of Imre Lakatos,* edited by Robert Cohen et al. Boston: D. Reidel, 1976.

Váli, Ferenc. *Rift and Revolt in Hungary.* Cambridge, Mass.: Harvard University Press, 1969.

Verso-New Left Books, eds. *The Value Controversy.* London: Verso, 1981.

Wartofsky, Marx. "The Relation between Philosophy of Science and History of Science." In *Essays in Memory of Imre Lakatos,* edited by Robert Cohen et al. Boston: D. Reidel, 1976.

Watkins, John. "Against 'Normal Science.' " In *Criticism and the Growth of Knowledge,* edited by Imre Lakatos and Alan Musgrave. Reprint, New York: Cambridge University Press, 1970.

Wechsler, James. *The Age of Suspicion.* New York: Random House, 1953.

Whitaker, Andrew. *Einstein, Bohr, and the Quantum Dilemma.* New York: Cambridge University Press, 1996.

White, Hayden. *Metahistory: The Historical Imagination in Nineteenth-Century Europe.* Baltimore, Md.: Johns Hopkins University Press, 1973.

——. *Tropics of Discourse: Essays in Cultural Criticism.* Baltimore, Md.: Johns Hopkins University Press, 1978.

Wolff, Robert Paul. *Understanding Marx: A Reconstruction and Critique of "Capital."* Princeton, N.J.: Princeton University Press, 1984.

Worrall, John. "Imre Lakatos (1922–1974): Philosopher of Mathematics and Philosopher of Science." In *Essays in Memory of Imre Lakatos,* edited by Robert Cohen et al. Boston: D. Reidel, 1976.

——. "Thomas Young and the 'Refutation' of Newtonian Optics: A Case-Study in the Interaction of Philosophy of Science and History of Science." In *Method and Appraisal in the Physical Sciences: The Critical Background to Science, 1850–1900,* edited by Colin Howson. New York: Cambridge University Press, 1976.

Young, John. *Totalitarian Language.* Charlottesville: University Press of Virginia, 1991.

Youschkevitch, A. P. "The Concept of Function up to the Middle of the Nineteenth Century." *Archive for History of Exact Sciences* 16 (1976): 37–85.

Zahar, Elie. "Why Did Einstein's Programme Supersede Lorentz's?" In *Method and Appraisal in the Physical Sciences: The Critical Background to Science, 1850–1900,* edited by Colin Howson. New York: Cambridge University Press, 1976.

Zimán, Mária Izsák. *A Memorial of Words for Eva: 1925–1944.* Unpublished translation by Hajnal Csatorday of *Betúsirko Évának: 1925–1944.* Ráánáná, Israel: Privately published, 1989.

Index

Adorno, Theodor, 198, 305, 310, 319n5, 356n52
Analysis-synthesis, 131–33, 140, 334n3
Anglo-American philosophy, 2, 3, 12, 14, 17, 147, 149, 155, 167, 197, 292, 258, 312
Aristotle, 132, 182
Arnauld, Antoine, 97, 104

Beria, Lavrenti, 266
Bildung (bildungsroman), 29–37, 42, 67, 108, 136, 138, 153, 155, 158, 292, 296, 315; biogenetic law and, 4, 27–31, 36, 68, 139, 140–41, 323n19; *Phenomenology of Spirit* and, 30–39; *Proofs and Refutations* and, 14, 27–30, 35, 38–39, 292. *See also* Caricature; Goethe, Johann Wolfgang von: *Wilhelm Meister*

Caricature (gestalt), 4, 29, 31–32, 34, 36–37, 139, 160, 194, 213, 218, 224, 231, 278, 293, 298, 301, 310, 312, 314, 344n2, 246n16. *See also* Historiography: falsification
Carnap, Rudolf, 305, 321n23, 355n39
Cartwright, Nancy, 341n14
Clark, Peter, 173–76, 178
Classical political economy. *See* Labor theory of value
Colletti, Lucio, 235–36, 342n16, 348n22
Concept stretching. *See* Conceptual novelty
Conceptual novelty, 25, 50–51, 68–76, 188, 327n23; concept stretching, 62–65, 85–88, 108–9, 126. *See also* Method of proofs and refutations: lemma incorporation
Contradiction, 16, 101, 154, 155, 162–63, 165–66, 189–90, 192, 196–201, 210, 212, 222, 297, 341n14; Hegel and, 101, 190–91, 197–99, 201; Popper and, 12, 158, 188–90. *See also* Hegel, G. W. F.: Aufhebung; Science: quantum theory
Covert ideas: Hungary and, 299, 301, 307, 315, 356n44; Lakatos and, 2, 5, 6, 140–42, 144, 300–301, 258, 313; Lukács and, 299, 303, 307, 310–11; Marx and, 7, 17, 139–40, 316; in mathematics, 131, 133; and reversal of discovery and justification, 138–39, 335n11. *See also* Analysis-synthesis
Criticism: as exoteric, 18, 212, 224, 315; Hegel's of Kant, 190–91, 198–99, 229–30; historical character of, 164; Lakatos's of Popper, 1, 5, 154, 159–64, 166, 215, 232, 292, 301, 314–16; Marx's of Hegel, 210–12, 216, 301, 309, 315. *See also* Phenomenological method; Research programmes

Davidson, Donald, 335n1
Descartes, René, 28, 91, 131, 136
Dialectic, 12, 17, 21, 28, 61, 65, 231, 269, 298, 321n31; Hegel and, 191, 229–33, 342n19; Kant and, 229–31; Marx and, 17, 233–34; self-criticism, 5, 8, 228, 275–76, 278, 280–83, 285, 296–97,

Dialectic (*cont.*)
300–301, 303–4, 309, 311; subject-object, 61–65. *See also* Contradiction; Hungarian Revolution: inverted world; Phenomenological method; Research programmes: changing logic of; Skepticism
Dialectical materialism, 5, 18, 198, 267, 269, 342n16
Duhem, Pierre, 160–61, 177, 196, 323n19
Dulles, John Foster, 268

Economics. *See* Labor theory of value; Science: economics
Economists: Blaug, Mark, 318n5, 350n20, 351n22; Bortkiewicz, Ladislaus von, 251; Dobb, Maurice, 255; Gillman, Joseph, 253; Meek, Ronald, 252–53, 349n5; Ricardo, David, 17–18, 237–50; Samuelson, Paul, 255; Smith, Adam, 17–18, 237–39, 242, 250, 348n3, 350n11; Sraffa, Piero, 241, 254; Steedman, Ian, 352n40
Eichmann, Adolph, 273
Elites, 223–25, 292, 296, 300, 307, 312. *See also* Hungarian Revolution: elites
Engels, Friedrich, 18, 30–31, 140
Error, 9–14, 34, 161; Goethe on, 32–33; Hegel on, 10–11, 34, 42, 66; in mathematics, 23, 54–55, 68–69, 71, 328n1; Popper on, 10, 159. *See also* Bildung; Contradiction; Falsification; Historiography: falsification of history; Mathematics: counterexamples
Euclidean style, 30, 35, 40, 43, 133–38, 153, 296, 315. *See also* Mathematics: formalization

Falsification (refutation): of methodologies of science, 161–62, 176, 212, 218, 222, 337n7; as negative, 159; in science, 158, 160–65, 189–90, 197, 207–8, 212, 217, 247, 253. *See also* Error
Feyerabend, Paul, 2, 6, 25, 332n3, 147, 151, 155, 307, 336n8, 339n30; *Against Method,* 149–50, 257, 318n5, 320n10, 335n2, 336n3; on Copernicus, 182, 339n31; on language, 144; on Newton, 225–26; skepticism of, 150, 330n22
Footnotes, 5, 13, 28–29, 38, 211, 213–14, 218, 291, 301, 318n5, 344n36

Gadamer, Hans-Georg, 323n19
Galison, Peter, 336n10, 343n23
Goethe, Johann Wolfgang von, 21, 292, 297; *Faust,* 1, 19, 38–39, 281, 309, 312; Hegel and, 342n19; *Wilhelm Meister,* 31–33, 38–39, 145, 230
Gomułka, Władysław, 270
Grattan-Guinness, Ivor, 71, 327n20

Habermas, Jürgen, 8, 202, 234, 298
Hacking, Ian, 6, 9, 202, 320n11, 343n25
Harris, H. S., 38
Haym, Rudolf, 37–38
Hegel, G. W. F.: Aufhebung (aufheben), 66, 161, 235–36, 298; history of philosophy and, 157, 336n3; inverted world, 269, 277, 298, 301–3, 307, 356n48; language and, 67, 85, 142–44, 219, 296, 290, 327n18, 335n14; logicism of, 216; mathematics and, 66–67, 323n19, 326n17; pedagogy and, 8, 30, 36–37; *Philosophy of Right,* 3, 4, 210; *Science of Logic,* 100, 232–33; self-consciousness, 102–3; skepticism and, 91–92, 329n7, 329n11. *See also* Bildung; Contradiction; Dialectic; Historiography; Hungarian Revolution: inverted world; Phenomenological method; *Phenomenology of Spirit*
Hegelian-Marxism, 2, 4, 6, 202, 269, 292, 296, 300, 308, 310, 314
Heraclitus, 140
Heuristic, 12, 28, 35, 40, 47–48, 94, 117, 126, 133, 228, 233, 297, 324n26. *See also* Analysis-synthesis; Research programmes: positive heuristic
Hindsight, 16, 155, 164, 190–91, 203, 207–8. *See also* Hegel, G. W. F.: *Philosophy of Right:* Science: Michelson-Morley experiments

Hirosige, Tetu, 188
Historicization: Hegel's of philosophy, 11, 157; Lakatos's of Popper, 12, 15, 158, 164, 244; Lukács's of Marxism: 16, 228; Marx's of political economy, 6, 140, 158, 235
Historiography: false consciousness and, 4, 225–26, 345n9; falsification of history, 4–5, 38, 125, 137, 213–14, 223, 225, 227, 291, 293, 298, 300, 302, 344n1, 353n7; Hegel and, 9, 16, 218–20, 324n21; historical phasing, 37–39, 195–96, 203, 226; imperfections in, 167; internal and external history, 27, 214–15, 218, 220–21, 286, 291–93, 300, 302, 345n7, 347n7; as normative, 151–53, 159, 211–12, 214, 217, 220, 277, 293, 344n36; *Proofs and Refutations* and, 24, 27–30, 56, 153, 159, 211, 213, 231, 265; rational reconstruction, 24, 29, 35, 100, 102, 154, 211, 213, 291, 293, 298, 301–2, 315–16, 324n24; as theory-laden, 29, 215–17, 219, 302, 316, 343n25. *See also* Caricature; Hindsight; Hungary: falsification; judgments using; *Phenomenology of Spirit;* Research programmes: normative role
Holton, Gerald, 223–24, 344n22
Hume, David, 91, 96–97, 104, 109, 120–21, 330n26
Hungarian Revolution (of 1956): compromise and, 4, 282, 304, 308, 310, 313–14; elites and, 270–74, 279, 280–82, 285, 289, 295–96, 301, 316, 352n5, 352n6; genesis of, 265–86; heroes of, 260–61, 285, 303, 311, 316; inverted world and, 3, 269, 293, 296–7, 310; language in, 272, 275–76, 281–82, 285, 296, 298, 308, 355n40; personality cult (*see* Elites); Petőfi Circle, 259, 283, 290–91, 296, 258, 304, 306, 356n40; role of truth in, 269, 272–73, 277–78, 279, 281–82, 284–86, 293, 303, 308, 310; writers and, 272–82, 284, 286, 291, 293, 296, 300–301, 303, 311. *See also* Covert ideas; Dialectic: self-criticism; Hungarians; Hungary
Hungarians: Aczél, Tamás, 269, 276, 303, 345n10; Ady, Endre, 272, 278; Déry, Tibor, 276–78, 280, 283, 285, 292–93, 303, 295, 306, 310; Farkas, Mihály, 266–67, 278–79; Farkas, Vladimir, 268; Fehér, Ferenc, 288; Gerő, Ernő, 260, 266–67, 278–79, 281, 283; Gimes, Miklós, 281; Hegedüs, András, 281; Horthy, Miklós, 266, 268, 276; Izsák, Éva, 288–89, 291; Jószef, Attila, 272; Kádár, János, 259, 268, 280, 282, 285, 299, 304, 308; Karácsony, Sándor, 287, 354n24; Keresztury, Dezső, 289; Konrád, Georg, 285; Köpeczi, Béla, 288; Kossuth, Lajos, 278; Kun, Béla, 265–69; Kutrovátz, Gábor, 287; Márkus, István, 288; Méray, Tibor, 269, 276, 303; Mérei, Ferenc, 288; Mindszenty, Jószef, 267; Münnich, Ferenc, 285; Nagy, Imre, 260, 278–81, 283, 285, 290, 303–4, 310, 355n38; Péter, Gabor, 266, 303; Petőfi, Sandor, 272, 283, 303; Polányi, Michael, 312; Rajk, Julia, 283–84; Rajk, László, 267–68, 271, 274–75, 280–84, 292–93, 304, 306; Rákosi, Mátyás, 259–60, 265–74, 278–80, 282–84, 286, 288–89, 293, 296, 298, 303, 308; Révai, Jószef, 258, 266–68, 274–78, 282–83, 289–91, 292–93, 303, 306; Révész, Éva, 287–88; Ropolyi, László, 287; Rudas, László, 274–75; Soros, George, 346n10; Szabó, Árpád, 12, 21, 90, 287, 290, 324n1, 332n4; Szász, Béla, 271; Vasvari, Pal, 283; Weisz, Alfonz, 288; Zimán, Mária, 291, 355n36. *See also* Lakatos, Imre; Lukács, Georg
Hungary: falsification in, 266–74, 279–80, 284, 285–86, 293, 296–97; Stalinism in, 265–78

Irrationalism: Hegel on, 39–43; Hungary and, 3, 299, 302; Lakatos on, 43; of Lakatos, 224, 312–13; Lukács on, 41;

Irrationalism (*cont.*)
of Lukács, 305–6, 311; mathematics and, 134–38; science and, 197. *See also* Lukács, Georg: *DR*; Elites; Euclidean style; Hungarian Revolution: inverted world; Schelling, Friedrich

Jammer, Max, 193

Kant, Immanuel, 41, 57, 91, 189–90, 198–202, 229–31, 341n4, 342n23
Kolakowski, Leszek, 253, 306, 310, 357n60
Khrushchev, Nikita, 281–83, 352n5
Kuhn, Thomas, 2, 154–55, 166, 218, 312, 338n15, 341n7; on Copernicus, 179–81, 339n32; on scientific discovery, 339n22; *SSR*, 1, 147–49, 152–53, 167, 257, 337n13

Labor theory of value: algebra of, 243, 249–50, 351n24; criticisms of, 251–55, 352n40; Marx's social theory and, 244, 240, 349n6; predictions and, 252–53; as research programme hard core, 239–40, 246, 351n22; transformation problem, 249–52. *See also* Marx: *TSV*; Economists; Science: economics
Lakatos, Imre: as educator, 8, 13, 19, 314; Hegelian allusions, 4, 27, 65, 210; Hegelianism of, 1–20, 294–95; in Hungary, 286–92, 313; language and, 86, 142, 212, 219, 296, 290; philosophy of mathematics (see *Proofs and Refutatins*); philosophy of science (*see* Research programmes). *See also* Covert ideas; Criticism; Footnotes; Historiography; Irrationalism; Lukács, Georg; Popper, Karl; Rationality
Lassalle, Ferdinand, 140
Lemma-incorporation. *See* Method of proofs and refutations
Lenin, Vladimir, 18, 273
Lichtheim, George, 305
Logical positivism, 157, 159, 160–61, 173, 175–77
Long, Jancis, 353n20, 355n36
Lukács, Georg, 39, 41, 292, 297, 300, 301, 319n5, 321n28, 324n24, 345n7, 353n9, 353n10, 357n60, 365n51; *DR*, 303–14; *H&CC*, 7–8, 16–17, 19, 224, 228, 311; Hegel and, 274, 276; in Hungary, 265, 274–77, 283, 285; interpretation of Marx, 6–7, 140, 142, 211, 236, 257, 300, 320n12, 320n13; Lakatos and, 287–88, 312–16. *See also* Hungarian Revolution; Irrationalism

Mann, Thomas, 263, 278, 311, 319n5
Marx, Karl: alienation, 302, 335n15, 356n49; *Capital,* 139–40, 158, 233, 237, 240, 244, 248, 251, 293, 335n11, 349n7, 350n11; *Critique of Hegel's Philosophy of Right,* 212; *Economic-Philosophical Manuscripts,* 7, 320n12; *Eighteenth Brumaire of Louis Bonaparte,* 323n17; *The German Ideology,* 216; *Grundrisse,* 7, 139–40; Hegel and, 142, 210; historiography and, 139, 215, 244; Kant and, 202, 342n23; *Theses on Feuerbach,* 210; *TSV*, 244–47. *See also* Covert ideas; Elites; Footnotes; Hegelian-Marxism; Historiography: historical phasing; Labor theory of value; Lukács, Georg: interpretation of Marx
Mathematicians: Abel, Niels, 28, 55; Archimedes, 131; Baire, René, 75; Bernays, Paul, 124, 334n20; Bolzano, Bernard, 89, 113; Boolos, George, 128, 334n22; Borel, …mile, 75; Bourbaki, Nicolas, 137, 335n8; Brouwer, L. E. J., 28, 87; Cantor, Georg, 71, 73–75, 81, 133, 327n23; Cauchy, Augustin, 28–29, 45, 53–7, 63, 68–70, 115, 127, 137, 321n23, 325n2; Cohen, Paul, 81, 122–23, 332n11; Dini, Ulisse, 71; Dirichlet, Peter Lejeune-, 54–6, 63, 68–72, 134, 137, 327n21; Euclid, 137; Euler, Leonhard, 56, 76; Feferman, Solomon, 126–28, 333n17, 325n2, 328n26; Fourier, Jean Baptiste, 54, 68, 77; Gauss, Carl,

60, 69; Gödel, Kurt, 83, 106, 136–38, 332n9, 334n6, 333n20; Hadamard, Jacques, 75–76; Hankel, Hermann, 62, 73–74, 327n21; Hardy, George, 137; Heine, Heinrich, 72; Henkin, Leon, 125; Hilbert, David, 83, 124, 129, 334n20; Kreisel, Georg, 126, 331n3, 333n16; Kripke, Saul, 128; Lagrange, Joseph Louis, 60; Lebesgue, Henri, 75–76, 80, 327n21; Lipschitz, Rudolf, 73; Löb, Martin, 124–25; Peano, Giuseppe, 80; Poincaré, Henri, 27, 61, 74, 111, 113–14, 124, 137, 185–86, 325n4; Poisson, Denis, 54; Póly, George, 5, 17, 27, 324n26; Ramanujan, Srinvasa, 138; Rényi, Alfréd, 290; Riemann, Bernhard, 54, 68, 70–71, 74–75, 326n13; Rosser, Barkley, 125, 128–29; Russell Bertrand, 81, 106, 115; Seidel, Philip, 54–56, 325n5, 326n10; Skolem, Thoralf, 332n5; Smith, H. J. S., 71; Szegö, George, 324n26; Tarski, Alfred, 89; Wallis, John, 131; Weierstrass, Karl, 29, 54, 56, 59, 68–69, 90, 115; Whitehead, Alfred North, 106, 115, 152; Zermelo, Ernst, 75

Mathematics, 76–83; choice functions (axiom of choice), 75–76, 137, 327n25; continuity, 54–56, 63, 70, 76, 79; continuum hypothesis, 74, 82, 117, 123, 334n6; counterexamples, 47–48, 54–61, 68–69, 71, 86, 112, 114, 135; dominant theories, 26, 111–12, 114; extensional and intensional, 125–28, 333n17, 333n20; first-order logic, 82–83; formalization, 25–27, 49, 82–83, 87–90, 109, 111–15, 129–30, 325n6, 327n19, 331n3; functions, 24, 56, 62–63, 71, 75–78; Hilbert-Bernays-Löb conditinos, 124, 128, 135; historical status of, 14, 23, 25–26, 108, 124, 128–30; infinite series, 54, 55, 69, 77–78, 325n6; infinitesimals, 57, 76, 326n12; infinity (ordinals and cardinals), 73–74, 81–82, 133; integration (integral), 69–70, 72, 75, 78; logic, 15, 26, 49–50; measure, 75, 78, 327n24; Peano Arithmetic, 80, 82, 116, 118, 122, 127–28; polyhedra, 24, 45–47, 57, 64, 70, 111–13, 132; set theory, 74–77, 80–82, 106, 117–18, 327n24, 331n35; translation (*see* Mathematics: formalization); trigonometric and Fourier series, 24, 77–79; undecidability, 93, 116, 118, 120–21, 308. *See also* Hegel, G. W. F.: mathematics and; Mathematicians; Method of proofs and refutations; *Proofs and Refutations;* Theorems

Method of proofs and refutations: early nineteenth century and, 23, 51, 53–56; formalism and, 49–50, 52–53, 59–60, 63, 65–66, 111; later nineteenth century and, 57, 59–60, 68–75; lemma-incorporation in, 48–49, 58–60, 85, 133–34; monsters, 62, 66, 68, 70, 74–75, 90, 126–28, 130, 326n13; skepticism and, 105, 108–9. *See also* Mathematics; Phenomenological method; *Proofs and Refutations*

Mill, John Stuart, 151

Montaigne, Michel de, 96, 120

Musgrave, Alan, 1, 170–71, 258, 337n11, 339n22

Nicolaus, Martin, 140

Order of discovery and presentation. *See* Analysis-synthesis

Orwell, George, 153, 290, 298–99, 301–2

Pappus of Alexandria, 131–33

Pascal, Blaise, 108–9

Pedagogy, 8, 17, 32, 297; *Proofs and Refutations* and, 35–36, 40, 138, 306, 313. *See also* Bildung; Heuristic; *Phenomenology of Spirit:* pedagogy

Phenomenological method, 11, 297; historiography and, 52, 331n30; lemma-incorporation and, 51–52, 58–60, 64–68; skepticism and, 15, 39, 90–92, 97–104

Phenomenology of Spirit, 4, 9, 15, 51–52, 67, 216, 293, 313; historiography of, 30, 34–35; introduction to, 10–11; pedagogy and, 9, 30, 34, 36–37, 40–42, 102, 136, 307; preface to, 30, 42–43, 139, 306. *See also* Bildung; Phenomenological method; Skepticism
Plato, 21, 103, 122–23, 332n4
Popkin, Richard, 95, 97–103
Popper, Karl, 23, 34, 51, 147, 155, 158, 161, 213, 222, 290–91, 315; on crucial experiments, 12, 163, 190, 204, 208, 210, 212, 223; on Hegel, 141, 319n8, 335n13, 341n2; history of science and, 12, 13, 211, 221; Lakatos and, 1, 5, 8; language and, 3, 142–43, 198, 286, 290; as like Kant, 199, 229; logic and, 158, 164, 341n3; *LSCD*, 1, 158, 203, 257, 347n1; *The Open Society and Its Enemies,* 225, 257, 318n4, 320n8, 324n28, 345n10, 350n9, 356n42. *See also* Contradiction; Criticism; Error; Hindsight
Postone, Moishe, 349n6
Progress: mathematical, 64–68, 106–7, 142; as negative, 10; scientific, 153, 154–55, 159, 166, 309, 341n2. *See also* Research programmes: judgments using
Proofs and Refutations (*P&R*). *See* Analysis-synthesis; Bildung; Conceptual novelty; Euclidean style; Heuristic; Historiography; Footnotes; Mathematics; Method of proofs and refutations; Pedagogy; *Phenomenology of Spirit;* Theorems
Protagoras, 103, 122–23, 314

Quine, W. V. O., 26, 161, 196–97, 312
Quotation(s), 5, 9, 12, 229, 316, 318n5

Rationality (reason), 296–97, 299, 308, 310, 312; as process, 13, 16, 18, 21, 36, 40, 42, 49–50, 67, 104, 141, 201; truth and falsity combined in, 49–50. *See also* Bildung; Irrationalism; Kant
Ratinal reconstruction. *See* Historiography
Ravetz, J. R., 326n13, 346n11
Research programmes, methodology of scientific: ad hoc theories and, 165, 190, 194, 246, 337n12; applications of, 167–88, 337n10; belief in, 201; changing logic of, 155, 227–33, 265, 292, 347n1; as code of honesty, 160, 190, 223, 247, 350n17; defined, 164–66; hard core of, 165, 168, 170, 174–75, 180, 183–84, 187, 189, 191–92, 194, 210, 239, 241, 247–48, 342n14; historical status of, 202, 221, 229, 231; historiographical method and, 13, 15–17, 166, 214, 217, 227–28, 300, 337n11, 347n2; instant rationality and, 204, 208, 210; judgments using (progressive or degenerating), 166, 169, 173, 177–78, 180, 182, 185, 195, 211, 222, 224, 240, 252, 300, 345n5, 346n11; normative role, 4, 9, 158–59, 161, 196, 211–12, 215–16, 222, 344n37, 336n8; novel facts, 180, 336n1, 341n13; positive heuristic, 165–66, 168, 170, 176, 180, 184, 186–88, 192, 194, 222, 250, 254, 340n44; problemshifts, 163–66, 195, 201, 246; as a synthesis, 157, 159; theory-laden observation and, 160, 163, 189–90, 337n6; weaknesses of, 178, 182, 187–88, 204, 211, 346n11. *See also* Labor theory of value; Science: astronomy, Michelson-Morley experiments, quantum theory, relativity, theories of combustion, theories of heat, theories of light

Schelling, Friedrich, 40–43, 136, 138, 142, 144, 296, 306–7
Schiller, Ferdinand, 31, 33, 35
Schlegel, Friedrich, 145, 230
Schmidt, Alfred, 198, 202, 343n23
Science: astronomy, 178–82; atomic spectra, 191, 193–96; atomism, 173–74, 176; Brownian motion, 173, 177–78, 339n30; changing methods of, 16,

155, 202, 231; classical physics, 163, 165, 183–84, 186, 192–93, 209; economics, 17–19, 235–36, 245–47, 291, 297, 300; electromagnetism, 162, 165, 183–84, 192–93, 203; ether, 165, 183–84, 186; formalization of, 197, 342n2; Michelson-Morley experiments, 154, 166, 185, 191, 203–12; particle physics, 223, 346n11; quantum theory, 165, 166, 191–96; relativity (special and general), 163, 165, 182–88; studies, 346n17; theories of combustion, 162, 169–73; theories of heat, 162, 173–78, 187; theories of light, 162, 165, 167–69. *See also* Contradiction; Hindsight; Historiography; Labor theory of value; Research programmes

Scientists: Balmer, Johann, 195; Bayen, Pierre, 171; Bessel, Friedrich, 179; Black, Joseph, 171; Bohr, Niels, 191–96, 210, 220; Boltzmann, Ludwig, 173–74, 176, 178; Brahe, Tyco, 179, 181; Brown, Robert, 177; Bucherer, Alfred, 185; Cavendish, Henry, 170–72; Clausius, Rudolph, 173; Compton, Arthur, 194; Copernicus, Nicolas, 179–82, 188; Cuvier, G., 31; Davy, Humphrey, 172; Einstein, Albert, 176–78, 182–88, 193–94, 203, 210; Fitzgerald, George, 185; Fizeau, Hippolyte, 206; Fowler, Alfred, 196; Fresnel, Augustin, 53, 169, 203–9, 338n18, 338n19; Galileo, 131, 149, 178–79, 181–82, 220; Gibbs, Johann, 174; Haeckel, Ernst, 27, 30; Hertz, Heinrich, 184; Huyghens, Christian, 168–69; Kaufmann, Walter, 185; Kepler, Johannes, 179, 81; Kielmeyer, Karl, 31; Kirwan, Richard, 171; Kramers, Hendrik, 194; Lavoisier, Antoine, 170–73; Lord Kelvin (W. Thomson), 176; Lorentz, Hendrik, 183–88, 205–7, 209; Loschmidt, Joseph, 176; Lysenko, Trofim, 221, 290, 306; Mach, Ernest, 175, 177–78, 184, 209; Maxwell, James, 204; Michelson, Albert, 203–12; Miller, Dayton, 207, 343n29; Minkowski, Hermann, 187; Morley, Edward, 203–12; Newton, Isaac, 131, 168–69, 199, 225–26; Ostwald, Wilhelm, 177; Perrin, Jean, 173, 177–78; Pickering, Charles, 195; Planck, Max, 174, 176, 178, 186, 193; Priestly, Joseph, 170–73; Prout, William, 214; Ptolemy, 179–81, 188; Shankland, Robert, 207; Sommerfeld, Arnold, 194; Stahl, Georg, 170; Young, Thomas, 167–69, 338n17, 338n18

Skepticism (Pyrrhonism), 28, 39, 52–53, 60, 88–89, 141, 149, 155, 302–3, 315–16; ancient vs. modern, 91–92, 329n7; appearance, 15, 100–101, 104–5, 119, 330n17; epoche, 94; goals of, 94, 96, 98, 101–2, 108–9; historiography of, 328n5; immanent critique, 95, 99–100, 105–6, 119–20, 129, 283, 329n12, 355n39; isostheneia, 93–94, 101, 107, 118–20, 123, 314; mitigated, 39, 89, 105, 120–21, 123, 144, 155, 315, 330n26, 330n36; peritrope (self-application), 16, 94, 95–96, 103–4, 121–23, 227–29, 230–31, 276, 296–97, 300, 313, 331n32; problem of the criterion, 93–94, 98, 101, 106–7, 117–19, 123; Pyrrhonists, 91–92, 94, 328n6; trope from relation, 98–99, 101, 107. *See also* Concept stretching; Dialectic; Hume, David; Phenomenological method; Protagoras; Theorems: Gödel's incompleteness

Spinoza, Baruch, 140, 355n39

Stalin, Joseph, 266, 270–71, 278–79, 282, 293, 297–98, 304. *See also* Hungary: Stalinism in

Steiner, George, 305

Teleology, 33, 35

Theorem(s): Cauchy's false continuity, 54–55, 68–69; as conjectures, 45; Euler's on polyhedra, 28, 35, 45–48, 50, 57, 61, 63–64, 66, 111, 114, 124, 132, 137, 325n2; Gödel's incompleteness,

Theorem(s) (*cont.*)
25, 30, 49, 52–53, 83, 115–30, 135, 332n7, 331n32; Hankel's false, 71–72; on infinite series, 55, 69; Paris-Harrington, 332n8; on point sets, 69–72, 133; and proofs as thought-experiments, 45–46, 121–22, 128, 324n1; on trigonometric and Fourier series, 35, 54–55, 69–70, 72, 133, 327n23; unprovability of consistency, 116, 118, 121, 124–28, 135. *See also* Mathematics

Tito, 267, 280–81, 284
Toulmin, Stephen, 312
Trojan horse, 8, 310, 320n8, 320n10, 356n57

Wartofsky, Marx, 6
White, Hayden, 344n4
Wittgenstein, Ludwig, 142–43, 312
Worrall, John, 87, 168–69, 331n1

Zahar, Elie, 87, 179, 181, 184–85, 331n1, 336n1, 337n12

John Kadvany is a Principal at Policy and Decision Science, a management and environmental risk consulting firm. He has published on Lakatos, the philosophy of mathematics, risk, and environmental policy. He lives in Menlo Park, California, and can be reached at guisesofreason@aol.com.

Library of Congress Cataloging-in-Publication Data
Kadvany, John David.
Imre Lakatos and the guises of reason / John Kadvany.
p. cm. — (Science and cultural theory)
Includes bibliographical references and index.
ISBN 0-8223-2659-0 (cloth : alk. paper)
ISBN 0-8223-2649-3 (pbk. : alk. paper)
1. Science—Philosophy. 2. Lakatos, Imre. I. Title.
II. Series.
Q175.K12 2001 501—dc21 00-049107