A Historical Sociology of Neural Network Research

José Miguel Olazaran Rodríguez

(Mikel Olazaran)

PhD

University of Edinburgh

1991

To Txarete Ganboa, with love

Txarete Ganboari, maitasunez

I hereby declare that this thesis is the result of research and composition undertaken solely by myself.

J. M. Olazaran

◆ Abstract

It has been argued that science is generated and validated through processes of controversy, and that controversies are 'closed' through 'rhetorical' processes of 'enrolment of heterogeneous allies and resources.' It has also been argued that, once a controversy is closed, it is increasingly difficult for the 'losing' position to maintain the plausibility of its views, arguments, and interpretations (words like 'reification,' 'inertia,' and 'institutionalisation' have been used to refer to this).

Controversies have shaped neural network research throughout its history, from the 1950s to the 1980s. In this dissertation I analyse the history of neural network research using a 'controversy/rhetorical tactics/enrolment of allies and resources/closure' scheme. I claim that the result is a useful and powerful interpretation of the main developments of the evolution of neural network research. The neural network controversy is especially interesting because it was once (in the late 1960s) closed against neural networks, and twenty years later (in the late 1980s) it was reopened. The history of neural network research can be seen as the history of the closure and reopening of the neural network controversy.

◆ Acknowledgements

This PhD project was financially supported by a Basque Government (Eusko Jaurlaritza) grant for postgraduate research at Edinburgh University.

This project would have been impossible without the help of many people and friends.

There are two people whom I would like to thank *very especially*. One is Donald MacKenzie, my thesis supervisor. His advice, commitment, and encouragement throughout the stages of this project have been invaluable. The other one is Jesús-María Larrazabal (University of the Basque Country). He introduced me to the social studies of science, and he encouraged me to come to Edinburgh. He has constantly encouraged, supported, and advised me in my work over the years. *Eskerrik asko.*

Special thanks are due to Peter Dayan and David Willshaw (both from the University of Edinburgh) for their comments on a late version of the whole draft manuscript. I wish to express my gratitude to James Fleck and Alfonso Molina (both from the University of Edinburgh) too, for comments and advice on earlier drafts.

My research trip to the United States in the autumn of 1989 was a critical moment in this project. I am extremely grateful to those I interviewed (see list in appendix 2) for their time, attention, and interest in my work. I am also grateful to those who provided me with valuable information by letter (see appendix 3: personal communications by letter). My stay in the US would have been much more difficult without the help and friendship of Joel Kallich and Sue Jennings. They came to my rescue in many critical moments (the US is no place to live if your credit card is lost somewhere between the Basque Country and Boston!). Thanks are due to Everett Mendelsohn too, for

providing me with a place to work at the Department of the History of Science of Harvard University.

I would also like to thank the following individuals for encouraging and supporting me in my research: Josetxo Beriain (Public University of Navarre); Miguel A. Quintanilla (University of Salamanca), Jesús Ezquerro (University of the Basque Country), and Miguel Muñoz (University of Edinburgh).

Many thanks to my parents, Soledad and Josetxo, for their love and support, and for their respect for my work over the years.

Writing a PhD dissertation has an inevitable element of loneliness and monologue. Transition back to the dialogue of life with Txarete Ganboa was always a pleasure. *Mila esker, Txarete.*

Finally, thanks a lot to *all* my friends.

CONTENTS

◆ ONE

Introduction

In this dissertation I study the historical evolution of neural network research from a sociological point of view. Neural network research is an approach to the problem of building intelligent machines (artificial intelligence) and studying and modelling (computationally) perception and cognition (cognitive science).

Neural networks (also called artificial neural networks, connectionist networks, parallel distributed systems, and neural computing systems) are information-processing systems composed of many interconnected processing units (simplified neurons) which interact in a parallel fashion to produce a result or output. The massively parallel architecture of these systems is remarkably different from that of a conventional, von Neumann digital computer. Furthermore, neural network systems are not programmed, but 'trained.' Training a neural network in some classification task involves selecting a statistically representative sample of input/output pairs and an algorithm for adjusting the strengths (the weights) of the connections between processing units when the system does not produce the desired outputs.

The neural network approach also differs from the tradition which has dominated artificial intelligence (AI) and cognitive science in the last decades, namely the symbol-processing approach. Within the symbolic approach, intelligence and cognition are seen as processes of symbol manipulation and transformation. A symbol-processing AI system relies on its representational structures and on the possibility of applying structure-sensitive operations to those structures. Representational structures are manipulated and transformed according to certain rules and strategies (algorithms), and the resulting expression is the solution to a given problem.

Researchers expect neural networks to have considerable success in tasks not easily programmable within the rule-based symbol-processing approach such as pattern recognition and

speech recognition. The learning capabilities of neural networks may be especially important for that type of tasks

I approach the history of neural network research from a sociological point of view. The sociology of science studies the processes of production and validation of scientific knowledge. In the 1970s, emphasis in the sociology of science changed from the study of professionally-defined institutions, organisations, and communities involved in the production of science (such as research specialties and disciplines) to the study of the genesis and development of scientific knowledge itself. In the last two decades a significant number of case studies have shown the social character of the processes of generation and evaluation of scientific knowledge. In other words, social processes — both internal to science and involving the wider society — are at the basis of scientific activity.[1] Of course the term 'social' is not used here in a pejorative sense, but in an interpretative sense. To say that science is socially constructed and validated is not to say that science is 'ideological' or 'bad.' If the knowledge-generating and knowledge-assessing activities of science are social processes, then those processes should be studied from a sociological perspective.

As Harry Collins (1981a, 1983) pointed out, one useful starting point in the sociological study of science is the interpretative flexibility of scientific evidence (results, experiments). 'Interpretative flexibility' means that scientific evidence (results, experiments) can always, in principle, be interpreted in different ways. In other words, the interpretative flexibility of scientific data, experiments or results is, in principle,

---

[1] Barnes (1974) and Bloor (1976) are early statements of the foundations of these ideas. Barnes and Edge (1982) introduced and reviewed many central topics in the sociology of science, and offered a detailed bibliography of the research done in the area (including case studies) up until then. Shapin (1982) also reviewed some of the most important social studies of science. Collins (1981a, footnotes) contains a quite comprehensive bibliography. Knorr-Cetina and Mulkay (1983) contains papers representing some of the main recent approaches in the sociology of science. MacKenzie and Wajcman (1985) applied the 'social construction' approach to the study of technology, and this has been quite a fruitful line of research in recent years (see Bijker, Hughes, & Pinch, 1987).

unlimited. This basic principle of the sociology of science has been stated by several researchers.

> ". . . Limitless debates made possible by the unlimited interpretative flexibility of data . . ." (Collins, 1983, p. 95)
>
> ". . . In principle, all the assumptions that go into scientific arguments can be challenged." (Pinch, 1981, p. 146)
>
> "At the instrumental level, studies have shown that debate can be kept up as long as participants desire by challenging the particularities of any individual experiment. At the phenomenal level, studies have supported the general argument that there are, in principle, an arbitrary number of interpretations of any set of data." (Pickering, 1981, p. 65)
>
> ". . . No knowledge possesses absolute warrant, whether from logic, experiment, or practice. There are always grounds for challenging any knowledge claim." (MacKenzie, 1990, p. 10)

But interpretative flexibility is only the starting point. The problem is how interpretative flexibility is reduced *in practice*, that is how a particular interpretation comes to be accepted as compelling or superior in a given situation. Donald MacKenzie (ibid., p. 11) formulated this idea in the following terms:

> ". . . It is important, as far as possible, to investigate why a given technical reason was found compelling, when, abstractly, it could have been challenged; and to ask what counts as superiority and efficiency in particular circumstances."

Interpretative flexibility is eliminated (or reduced to practical levels) through the closure of controversies (Collins, 1981a; 1983). In controversies, different (and sometimes alternative) interpretations are put forward, and researchers confront each other's claims. Thus the interesting problem is to study the processes and mechanisms through which controversies — which in principle could always go on — are closed in practice. This controversy/closure scheme was used in a number of case

studies in the sociology of science (e.g. Collins, 1981c), and was later 'exported' to the sociology of technology (e.g. Bijker, Hughes, & Pinch, 1987; Pinch & Bijker, 1987). Bruno Latour (1987) developed it further to formulate a comprehensive approach to 'technoscience' as an heterogeneous network involving a variety of resources and 'actants' that are linked (or associated) together in processes of enrolment and enlisting (i.e. by controlling the 'behaviour' of those actants). Latour used the term 'technoscience' to encompass all the activities related to research, and to go beyond artificial boundaries such as the one between 'science' and 'technology.' This choice is particularly useful in the case of neural network research which is, by definition, an heterogeneous network including elements from 'science' (the study of the brain, perception, intelligence, and cognition . . .), and 'technology' (. . . using computational models, i.e. machines). Here I will use the term 'research' or 'scientific research' (in a general sense) instead of Latour's 'technoscience' term, but the sense in which I use those terms is not far from Latour's.

A crucial characteristic of scientific controversies is the use of rhetorical and tactical resources and arguments (Latour, 1987; Star, 1989a). Closure is not caused by the mere addition of scientific evidence, or by the 'technical superiority' (or 'superior rationality') of one of the positions. That superiority is the result of the closure of the controversy, and therefore it cannot be the cause.[2] Harry Collins argued that, if a debate is ever to be closed, rhetorical tactics must be employed:

> "Some 'non-scientific' tactics *must* be employed because the resources of experiment alone are insufficient." (Collins, 1985, p. 143)[3]

---

[2] Collins (1985, p. 106, fn. 6) expressed this idea in the following terms: ". . . The success of one party to a dispute of this sort *cannot* be explained by their superior grasp of the nature of the phenomenon under investigation. It is this that it is being discovered (determined) by the debate itself . . ." The idea is also one of Latour's (1987, p. 258) rules of method: "Since the settlement of a controversy is the cause of Nature's representation, not its consequence, we can never use this consequence, Nature, to explain how and why a controversy has been settled."

[3] (Collins, 1985, p. 152) is a similar statement.

> "Within the relativist programme we accept . . . that the scientific view belonging to both sides of a controversy can be defended indefinitely and that *even in the purest of sciences*, if debate is going to end, it must be brought to a close by some means not usually thought of as strictly *scientific* . . ." (Collins, 1983, p. 99)

(In section 3.1 I discuss why I do not like using terms like 'non-scientific' tactics, even if they are written in inverted commas.) Rhetorical tactics are always used in scientific controversies. They are elements of scientific discourse and practice (and, indeed, of discourse and practice in general). In other words, scientific discourse and practice are organised and constructed through the use of rhetorical arguments and resources. I do not use the term 'rhetorical tactics' in a pejorative sense. Quite the opposite: the 'move' in science is not from rhetoric to 'truth,' but from weaker rhetoric to stronger rhetoric, as Bruno Latour (1987) pointed out. Scientific discourse mobilises more allies and resources than (say) everyday discourse, and therefore in this sense it is more social than everyday discourse. Latour (1987, p. 62) enunciated his idea in the following terms:

> "We saw a literature [scientific literature] becoming more technical by bringing in more and more resources . . . We saw a dissident driven into isolation because of the number of elements the authors of scientific articles mustered on their side . . . The more technical and specialised a literature is, the more 'social' it becomes, since the *number of associations* necessary to drive readers out and force them into accepting a claim as a fact increase . . . If being isolated, besieged, and left without allies and supporters is not a social act, then nothing is. The distinction between the technical literature and the rest is not a natural boundary; it is a border created by the disproportionate amount of linkages, resources and allies locally available. This literature is so hard to read and analyse not because it escapes from all social links, but because it is *more* social than so-called normal social ties."

The result of the closure of a controversy is (depending on the nature of the debate) the replicability of an experiment, or the validity of some results or of some standards of scientific practice (models of problem/solutions). But accepting an interpretation of a result (or a set of results) means that other, contending interpretations of that same result are rejected. Supposing that the controversy is between two positions, closure of the controversy in favour of one of them brings about the rejection of the other one (of course this may be a matter of degrees of acceptance and rejection). Thus when a debate is closed there are 'winners' and 'losers.' One position wins when its mobilisation of resources, actants and allies cannot be contested by the opposing position, that is when it is successful in tipping the balance of power in its favour. When this happens, the 'losing' side either accepts their opponents' position as a 'fact,' and uses it as a 'back box' in its own research practice, or keeps working in a quite isolated way (far from the hot centres of scientific activity) hoping that one day it can launch a counter-attack and contest the *status quo* emerging from the first closure (but their opponents' view is accepted as correct in the mean time anyway). The closure of a controversy can be seen as the establishment of some relationships of power, and therefore it is not consequence of the intrinsic 'technical superiority' of one of the positions (again: that is what is being decided in the controversy).[4]

Once an interpretation has emerged as the 'correct' one after the closure of a controversy, time runs against the losers as the institutionalisation of the winning side gains momentum (terms like 'inertia' and 'reification' have also been used to refer to this phenomenon). It is increasingly difficult for the researchers who supported the losing side to show the plausibility of their arguments. As (postclosure) scientific activity normalises and

---

[4] This notion was formulated by B. Harvey (1981, p. 124) in terms of monopolisation of plausibility: "The suggestion is that the winning side does not posses truth, but rather that it has monopolized plausibility."

institutionalises, controversial episodes are quickly forgotten, and the winning view gains an increasing appearance of self-evident 'truth'.[5]

> "Even when it is pointed out that the viewpoint of the 'losers' is logically tenable, it is difficult for the reader to remain impartial in the face of the sheer weight of numbers in the 'winning' camp." (Harvey, 1981, p. 126)

Harry Collins (1975, pp. 94-95) used the analogy of the ship in a bottle to describe this process:

> ". . . Much of our knowledge seems so 'solid' as to require a justification in terms other than those which describe human actions . . . To speak figuratively, it is as though epistemologists are concerned with the characteristics of ships (knowledge) in bottles (validity) while living in a world where all ships are already in bottles with the glue dried and the strings cut. A ship *within* a bottle is a natural object in this world, and because there is no way to reverse this process, it is not easy to accept that the ship was ever just a bundle of sticks . . . [But] it is possible to perform a kind of phenomenological bracketing for ideas

---

[5] Thomas Kuhn made a similar point. After periods of change are over and a view has been accepted, the past is seen as developing linearly (cumulatively and naturally) toward the presently accepted view. ". . . To an extent unprecedented in other fields, both the layman's and the practitioner's knowledge of science is based on textbooks and a few other types of literature derived from them. Textbooks, however . . . have to be rewritten in the aftermath of each scientific revolution, and , once rewritten, they inevitably disguise not only the role but the very existence of the revolutions that produced them. . . Partly by selection and partly by distortion, the scientists of earlier ages are implicitly represented as having worked upon the same set of fixed problems and in accordance with the same set of fixed canons that the most recent revolution in scientific theory and method has made seem scientific. No wonder that textbooks and the historical tradition they imply have to be rewritten after each scientific revolution. And no wonder that, as they are rewritten, science once again comes to seem largely cumulative. Scientists are not, of course, the only group that tends to see its discipline's past developing linearly toward its present vantage. The temptation to write history backward is both omnipresent and perennial . . . The depreciation of historical facts 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 . . . The result is a persistent tendency to make the history of science look linear or cumulative, a tendency that even affects scientists looking back at their own research" (Kuhn, 1970, pp. 137-139). "There are no collections of 'readings' in the natural sciences. Nor are science students encouraged to read the historical classics of their fields — works in which they might discover other ways of regarding the problems discussed in their textbooks, but in which they would also meet problems, concepts, and standards of solution that their future professions have long since discarded and replaced" (Kuhn, 1977, pp. 228-229).

> and facts, by looking at them while they are being formed, before they have become 'set' as part of anyone's natural (scientific) world . . . This will generate a picture of science in which the figurative 'ships' are still being built by human actors, to be subsequently erected in their bottles by a trick invented and worked by human actors also."

Bruno Latour (1987) speaks about 'facts,' and 'black boxes' when he refers to this idea. What all this means is that in a sociological study of science one has to reconstruct the circumstances and processes in which a 'black box' was created, or a ship put into a bottle. This becomes more and more difficult as time passes by and the 'facts' emerging from the closure of a controversy are used as black boxes to construct more facts, and practices, 'forms of life,' and expectations which incorporate those facts as black boxes develop and institutionalise.

Furthermore, what all this talk about black boxes and ships in bottles also means is that the reopening of a black box (or the reopening of a bottle and the removal of the ship from it for revisions and changes) is an especially interesting case for a sociological study of science. A case in which the 'loser' (the side unable to contest the balance of power) in a controversy re-emerges sometime (say, two decades) later so as to reopen the controversy again (and change that balance of power) would be especially interesting for a sociological study of science.

Harry Collins sometimes makes the 'things could have been otherwise' point in his studies of mechanisms of closure. In his study of the controversies in gravitational radiation in the 1970s, Collins (1985, pp. 104-106) pointed out that, in accepting the electrostatic calibration measuring technique, J. Weber restricted the interpretative flexibility of gravitational radiation results, and chose not to argue in certain fronts which, in principle, were not entire implausible:

> "Weber in accepting electrostatic calibration chose not to argue on these fronts. My respondent's decision to open up the range of possibilities for calibration signals reveals

that such an argument might not have been entirely implausible."

Thus a case in which 'things that could have been otherwise' earlier *are actually* otherwise at a later point, in different circumstances, seems especially interesting from the point of view of the social processes at the basis of the knowledge-generating and knowledge-assessing activities of science. My claim is this thesis is that the history of neural network research is such a case in many important respects. I also claim that the 'controversy/closure/rhetorical tactics/enrolment of resources and allies' scheme — developed by authors like Harry Collins (1985), Bruno Latour (1987), and Susan Star (1989a) — that I have outlined in this introductory chapter provides a useful and powerful framework for reconstructing and interpreting the historical evolution of neural network research from the 1950s to the 1980s. In certain parts of this dissertation I also use some other 'interpretative tools.' For example, in section 4.2 I show that a 'metaphor scheme' (Barnes, 1974) is useful for studying certain recent developments in neural network research.

In the rest of this introduction I will first do a **chapter by chapter introduction** to the dissertation, and then I will make some 'methodological' comments and remarks about my research.

In **chapter two** ('Early Neural Networks') I look at neural network research in the 1950s and 1960s. In those years neural network research activity reached important levels. Up to the mid-1960s, neural networks was an important tradition of AI-like research. A number of neural network systems were studied and tested, and important implementation projects were carried out.

The origins of neural networks, as the origins of AI-like research in general, go back to the cybernetics movement of the 1940s and 1950s (section 2.1: 'Cybernetics and the origins of neural networks'). The brain/machine problem was a central one in

cybernetics, and a number of information-processing approaches to it were pursued. Symbol-processing AI and neural network research were only two of them. It is important to note that, in cybernetics, the brain/machine problem was often stated and studied using neural network terminology. At that time, neural network terminology was used in diverse (and even alternative) approaches to the brain/machine issue.

With the advent of the von Neumann computer in the 1950s, some researchers started to develop the brain/machine cybernetic theme in the brain/von Neumann computer direction (or, more accurately, mind/von Neumann computer direction). This tradition (the symbol-processing approach to AI), which started to gain increasing momentum as digital computers became more available throughout the 1960s, was opposed by early neural network researchers. The spokesman (using Latour's term) of the neural network position in this respect (as in many others) was Frank Rosenblatt, a 'cognitive systems' researcher from Cornell University. He developed many of the (now so popular) neural network ideas in the late 1950s and early 1960s. Rosenblatt's general conceptual framework can be studied from the point of view of the brain/computer issue (section 2.2: 'The computer and the brain'). The symbol-processing approach to AI and cognitive science uses the digital computer as its experimentation tool for studying intelligence and cognition and building intelligent machines. Thus it can be said that, for the symbolic approach, the von Neumann computer is a metaphor of cognition. For Rosenblatt, and the other early neural network researchers, it was the other way round. They aimed at using the brain itself as a metaphor for the construction of intelligent information-processing systems.

Big machines (in the literal sense of 'big') were built within this 'brain as metaphor' approach. The first important one — and the most famous one — was the Mark 1 perceptron built at Cornell Aeronautical Laboratory, an implementation of Rosenblatt's perceptron system (section 2.3: 'The perceptron'). One aspect of

the perceptron which created considerable interest in neural networks was its capacity for learning (i.e. for improving its performance significantly in) certain classification tasks. Rosenblatt's 'learning algorithm,' and a learning technique developed by Bernard Widrow and Marcian Hoff at Stanford University, were very important in early neural network research.

But the perceptron was not the only big project of early neural network research. Other projects and machines followed. Two very important developments in this respect were the Madaline and the Minos machines (section 2.4: 'The Madaline and Minos projects'). The Madaline was built by Bernard Widrow and his colleagues. Minos was built at Stanford Research Institute by Charles Rosen, A. E. Brain, and others. Research at both Widrow's laboratory and Stanford Research Institute shows that the neural network approach to AI was pursued and developed seriously by a significant number of researchers in the 1950s and 1960s.

At the same time, the emergence of the symbol-processing approach was gaining increasing momentum (section 2.5: 'The emergence of symbolic artificial intelligence'). These researchers were building and studying intelligent systems using tools and techniques which were rather different from those used by neural network researchers. But symbolic AI was not only rather different from neural networks. The opposition between the two approaches was always explicit and active. Furthermore, in several critical moments of the history of neural network research, that opposition became open controversy.

In **chapter three** I study one such critical moment, namely the controversy about Rosenblatt's perceptron machine and about neural networks in general (Three: 'The Perceptron Controversy'). The opposition between neural network and symbolic AI researchers became open (and sometimes not very diplomatic) controversy in the late 1950s and early 1960s (section 3.1: 'The heat of the controversy'). Rosenblatt's claims about his perceptron project were heavily contested by symbol-processing

researchers, and an open and at times bitter controversy developed. The rhetorical tactics used by both sides are a good indication of the extent of the controversy.

Early neural network researchers often recognised that their systems had important problems and limitations, and were aiming at building and studying more complex machines. However, by the mid-1960s, the perceptron controversy was affecting neural network projects seriously, and many researchers started to feel powerless to contest the arguments and criticism developed by researchers who were against the neural network approach. The problems of single-layer neural network machines were becoming increasingly apparent, and work in more complex systems (e.g. multilayer networks) was getting rather difficult (section 3.2: 'The crisis of early neural networks'). The crisis of early neural networks affected the three main projects differently. Widrow and his colleagues started to develop applications of their neural network techniques in the area of adaptive signal processing. The Stanford Research Institute researchers moved to symbol-processing AI and started an important robotics project. Rosenblatt and his colleagues acknowledged the problems and limitations of their machines, but continued to look for solutions within the neural network approach.

In the early 1960s — before the just mentioned crisis of neural network research — Marvin Minsky and Seymour Papert, two symbolic AI researchers from the Massachusetts Institute of Technology, decided to start a project which, if successful, would have a decisive effect in the outcome of the perceptron controversy. Their aim was to show clearly and decisively the limitations of Rosenblatt's perceptron (and similar neural network systems). They 're-enacted' (Latour, 1987) Rosenblatt's perceptron with a view to finding the 'flaws' and problems it contained. Minsky and Papert's main arguments against the perceptron in particular and neural network research in general were known by the mid-1960s, and had had a significant effect

on the crisis of neural networks before their (1969) study was published. Minsky and Papert's (1969) study (their famous 'Perceptrons' book) was the result of the above mentioned re-enacting process.

It is usually thought that Minsky and Papert (1969) study showed the limitations of perceptrons beyond doubt. That study is also usually interpreted as having showed that neural network research was not worth pursuing. In section 3.3 ('Interpretative flexibility') I show that this view was the result, and not the cause of the closure of the perceptron controversy. Minsky and Papert's (1969) results on the problems of the single-layer perceptron were open to interpretative flexibility. Even more open to interpretative flexibility were Minsky and Papert's comments on the capabilities of more complex perceptrons (multilayer perceptrons). Minsky and Papert's (1969) study was open to interpretative flexibility not only in principle, but (very importantly) in practice. Neural network researchers tried to exploit that interpretative flexibility in their favour.

Nevertheless, Minsky and Papert's (1969) arguments were widely interpreted as showing that the neural network approach was not worth pursuing (section 3.4: 'Closure'). Researchers from Rosenblatt's group tried to exploit the interpretative flexibility of Minsky and Papert's (1969) arguments in their own rhetoric, and insisted on the 'promising aspects' of more complex perceptrons. But this was not enough to contest Minsky and Papert's arguments and the conclusions which had been drawn from them. Rosenblatt and his colleagues were unable to enrol enough resources and allies to contest Minsky and Papert's 'last word effect' and maintain the plausibility of neural networks as an approach to AI. The controversy was over (and there were 'winners' and 'losers'). Researchers against the neural network position were successful in linking their criticism of neural networks with factors such as the emergence of symbol-processing AI and the development of von Neumann computer

technology. These were two important closure mechanisms in the perceptron controversy.

In the 1970s some researchers continued to work on neural networks, but they were far from the 'hot' centres of AI and cognitive science activity. This was much more so in the United States than in Europe. After the closure of the perceptron controversy, these researchers retreated from AI-like activity to more neuroscience and psychology-oriented research. The dominance of the symbol-processing approach continued over the years, until the situation started to change in the early 1980s. In **chapter four** ('New Connectionism') I study some developments of the early and mid-1980s which made the re-emergence of neural network research possible.

By the early 1980s the world of AI and related areas of research was changing significantly. There had been very considerable developments in digital computer technology, symbol-processing AI had reached a stage of commercialisation, funding was increasing after the Japanese Fifth generation project, and several parallel computing architectures were being studied and developed. Developments in the area of information-processing were going fast. In this context, a group of researchers, the so-called Parallel Distributed Processing (PDP) group started to link several elements with a view to bring neural networks back to the AI-cognitive science front (section 4.1: 'Parallel distributed processing'). Among the allies and factors that the PDP group intended to enrol were the neural network researchers of the 1970s, researchers who were working within the symbol-processing approach but who were having difficulties in studying and modelling certain cognitive processes, researchers in pattern recognition, speech recognition, and vision, and developments in parallel computing. The PDP researchers did an important work in bringing neural network research back to the AI arena, and they did so in an explicit and active manner.

In the mid-1980s, neural network researchers (a significant number of them belonging to the PDP group) developed important

scientific innovations. Some of the most important of these innovations can be studied within a 'metaphor and innovation scheme' (section 4.2: 'Networks with symmetric connections: metaphors and innovation in neural computing'). In 1982 John Hopfield, a physicist from the California Institute of Technology, developed a neural network system based on an analogy from statistical physics. This analogy between certain systems in statistical physics and neural networks was developed further by David Ackley, Geoffrey Hinton, and Terrence Sejnowski (the last two from the PDP group) (1985) in their Boltzmann machine network. The Boltzmann machine was the first successful solution to the problem of learning in multilayer neural networks, an important point of Minsky and Papert's arguments in 1969.

Developments started to happen fast from then onwards. Those developments are studied in **chapter five** ('Controversy Reopens'). Researchers from the PDP group were working on another type of multilayer system: the so-called back-propagation network. Some researchers had worked on the idea of back-propagation before, but they had found considerable resistance within AI (section 5.1: 'History of back-propagation'). The idea was not accepted widely until it was developed in 1986 by PDP researchers David Rumelhart, Geoffrey Hinton, and Ronald Williams within the neural network framework (section 5.2: 'Back-propagation: learning in multilayer perceptrons'). Back-propagation, a technique for learning in multilayer perceptron-like systems, was the element that finally precipitated the reopening of the neural network controversy. Almost twenty years after the closure of the perceptron controversy, neural network researchers were now in a position to contest Minsky and Papert's challenge and to force a revision of the balance of power which resulted from that closure. Minsky and Papert made a counter-attacking move, in which they insisted that the newly developed neural network techniques had important limitations, and that many of their (1969) conclusions still remained valid. But controversy had reopened, and Minsky and Papert's arguments

did not have too great an effect this time round. The emergence and growth of neural network research was well under way by then (section 5.3: 'The neural network explosion').

The future map of AI and cognitive science research has not been clearly defined yet. It is still very much a matter of negotiation and debate. The situation of debate created by the re-emergence of neural network research continues (section 5.4: 'Debate continues'), and the resulting balance of power will be important for years to come.

Finally, in **chapter six** I develop the conclusions of this sociological study of the history of neural network research.

Before I finish this introduction I will make a few 'methodological' comments about the development of my research project. One difficulty that I found (which can be seen as a consequence of the closure of the perceptron controversy) was the lack of information about early neural network research. Many of the details of that period were deeply buried, and I had to carry out a considerable 'digging' effort. Some of the researchers whom I interviewed in the United States (see list in appendix 2) were particularly helpful in this respect. In-depth historical studies of the evolution of neural network research have not been done as far as I know, and this made my work especially difficult at times.

The issue of the lack of historical studies of AI research in general was pointed out recently by A. E. Adam (1990, p. 233) in his "plea for more attention to the history of artificial intelligence and also for the adoption of the sociology of knowledge, which has been the main historical methodology in the history of science in recent years." The lack of historical studies of neural networks was more worrying for me than the lack of 'sociological histories.' After all, the 'sociology of scientific knowledge' approaches were developed in the 1970s and 1980s with a view to applying them to as many (and as different) particular cases and episodes as possible. But the lack

of historical studies of early neural computing research had also a positive side, namely that I could develop my own approach to that history more 'freely,' so to speak.

In order to develop a history of neural network research, I had to reduce the complexity of the problem: there were infinitely many developments and events which could be linked to the evolution of neural network research. I hope that I made an interpretatively powerful simplification, given the obvious limitations of time and resources. In particular, I have focused on the development of certain neural network systems (namely feedforward and symmetric networks) and on certain neural network learning techniques (namely supervised learning). I have also focused more on AI-cognitive science aspects of neural networks, rather than on neuroscience-oriented ones. However, the reader should not forget that there is a big variety of neural network systems and learning schemes (let alone applications), and that what I cover in this dissertation is not the whole neural network field. This applies even more to the field of symbolic AI. The evolution of symbol-processing AI has been a key factor throughout the history of neural network research, but looking at it in detail is out of the scope of this dissertation. Therefore I had to extract the most general defining characteristics of symbolic AI and ignore much of the variety of techniques, tools, and schemes employed within that approach.

But having said this, I think that this simplification 'pays' in terms of the interpretation of the most important developments of the history of neural network research. I also think that, by applying the 'controversy/closure/rhetorical tactics/enrolment of resources and allies' scheme (discussed earlier in this chapter) to the study of the history of neural network research, I have developed a useful and powerful interpretation of that history.

◆ TWO

Early Neural Networks

2.1 Cybernetics and the origins of neural network research

In this section I discuss the origins of neural network research in the cybernetics movement of the 1940s and 1950s. In particular, I look at the central cybernetics issue of the relationships between brain and machine (the 'brain/machine' problem). The importance of McCulloch and Pitt's (1943) 'A logical calculus of the ideas immanent in nervous activity' paper in the brain/machine issue of cybernetics is emphasised. McCulloch and Pitt's ideas were used and developed in several different directions. Cybernetics emphasised information-processing, and in this it was a challenge to behaviourism. But it was not a uniform challenge. I show that one of the main characteristics of cybernetics was the variety of approaches to information-processing that were developed.

In the 1950s researchers started to combine McCulloch and Pitts' formal neurons with Hebb's notions of 'cell assembly' and 'synaptic modification.' These were the first contributions to neural network research. I look at one of them in particular: a machine built by Marvin Minsky and Dean Edmonds. I conclude by pointing out that, during the mid- and late 1950s, two approaches to studying cognition and building intelligent machines were emerging from cybernetics: symbol-processing AI and neural network research.

The origins of neural networks, like the origins of artificial intelligence (AI) and cognitive science in general, go back to the cybernetics movement of the post World War II years.[6]

---

[6] Antecedents can be found earlier, e.g. in associationist psychology. James Anderson and Edward Rosenfeld (1988, pp. 1-3) emphasise the similarity between some ideas developed by 19th century American psychologist William James and certain notions in current neural network research.

Cybernetics was a 'movement' affecting diverse scientific specialties (from engineering to social sciences) rather than a clearly defined area of research.[7] This movement developed through interdisciplinary conferences (such as the ones sponsored by the Macy Foundation in the 1940s) and contacts. The study of the relationships between automatic machines and the nervous system, and particularly between the brain and automatic machines (what I call here the brain/machine problem) was a central theme in cybernetics. Norbert Wiener, a mathematician from the Massachusetts Institute of Technology (MIT) who had worked on the improvement of anti-aircraft artillery during the war, defined cybernetics in the following terms

> ". . . Cybernetics attempts to find the common elements in the functioning of automatic machines and of the human nervous system, and to develop a theory which will cover the entire field of control and communication in machines and in living organisms." (Wiener, 1948, p. 14)

In the early 1940s, Wiener and his colleagues J. Bigelow and A. Rosenblueth became interested in the analogy between engineering devices and the nervous system in terms of feedback and control.[8] Wiener emphasised the importance of communication and control (and therefore information) for feedback and control.

> "What distinguishes communication engineering from power engineering is that the main interest of the former is not the economy of energy but the accurate reproduction of a signal." (Wiener, 1948, p. 15)[9]

---

[7] For a sociological study of cybernetics as a scientific 'subject-complex' see (Apter, 1972).

[8] An example of this discussed by Wiener is the governor of Watt's steam engine. Steady speed (homeostasis) is maintained by the effect on an action being fed back to the mechanism controlling the action (Pratt, 1987, pp. 193-194).

[9] "Information is information, not matter or energy. No materialism which does not admit this can survive at the present day," said Wiener in the 1961 edition of his 1948 'Cybernetics, or control and Communication in the Animal and the Machine' book (as quoted by Gardner, 1985, p. 21).

Rosenblueth, Wiener, and Bigelow's (1943) work can be seen as an attempt to study mechanisms which could embody 'purpose' (Papert, 1965, p. xxiii).

Another important development of the early 1940s was British psychologist Kenneth Craik's (1943) book on 'The Nature of Explanation.' Craik's emphasis on symbolism is an antecedent of the symbol-processing paradigm which became dominant in artificial intelligence (AI) in the 1960s. According to Craik, the nervous system parallels, models external events, that is it develops an internal model (representational mechanism) of reality.

> "My hypothesis then is that thought models, or parallels, reality — that its essential feature is . . . symbolism, and that this symbolism is largely of the same kind as that which is familiar to us in mechanical devices which aid thought and calculation." (Craik, 1943, p. 57)[10]

Craik's idea became later central to AI and cognitive science, as it was pointed out by de Mey (1982, p. xv):

> ". . . The principle that any form of information processing, whether natural or artificial, requires a device that has in some way or another, an internal model or representation of the environment in which it operates . . . This principle can be retrieved, almost in its bare form, from Craik's 1943 PhD thesis."

At about the same time as the papers by Wiener and colleagues and Craik's thesis, Warren McCulloch and Walter Pitts (1943, 'A logical calculus of the ideas immanent in nervous activity') studied how formal logic could be realised by a physical substratum of neuron-like elements. McCulloch and Pitts' paper

---

[10] See also: ". . . There is evidence of [symbolisation] in our own sensory and central nervous systems; and the function of such symbolisation is plain. If the organism carries a 'small-scale model' of external reality and of its own possible actions within his head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilise the knowledge of past events in dealing with the present and future, and in every way to react in a much fuller, safer, and more competent manner to the emergencies which face it" (Craik, 1943, pp. 60-61).

was of central importance for cybernetics research (and not only neural network research) on the the brain/machine question. Warren McCulloch was first professor of psychiatry at University of Illinois-College of Medicine and Illinois Neuropsychiatric Institute, and later he went to MIT.[11] Walter Pitts was a logician who worked at the Department of Mathematics at MIT. In the 1930s British logician Alan Turing had developed the notion of a machine which was capable of carrying out any effectively defined computation, that is any symbolic manipulation task defined by a finite and explicit set of rules (or algorithm). McCulloch and Pitts' (1943) paper can be seen as an attempt to show how propositional logic (i.e. the manipulation of abstract symbols) could be realised by a network of simplified neuron-like elements. In other words, they tried to show how nervous activity could support (realise physically) symbolic computations.

One of the main assumptions made by McCulloch and Pitts in their definition of a formal or simplified neuron was the binary character of its output — an assumption that would be a basic feature of neural network architectures for many years to come.

> "Because of the 'all-or-none' character of nervous activity, neural events and the relations among them can be treated by means of propositional logic. It is found that the behavior of every net can be described in these terms . . . and that for any logical expression satisfying certain conditions, one can find a net behaving in the fashion it describes." (McCulloch & Pitts, 1943, p. 18)

In figure 2-1 below some types of McCulloch-Pitts neurons are included. Minsky's (1967, pp. 33) diagrammatic notation is used, because it is clearer than the one originally used by McCulloch and Pitts. McCulloch and Pitts networks have excitatory and inhibitory connections. It is important to note that the connections in these networks have fixed value. This value is assumed to be 1 for the excitatory connections in the diagrams

---

[11] Some of McCulloch's papers were published in (McCulloch, 1965; new edition 1988).

below. The action of inhibitory connections is absolute. If a neuron has an incoming inhibitory connection, and that connection is activated, the neuron will not fire. A neuron fires (sends output activation) if the sum of its inputs equals or exceeds its threshold value. The activity of the network occurs in discrete time delays. Each neuron updates its state once in each time delay.[12]

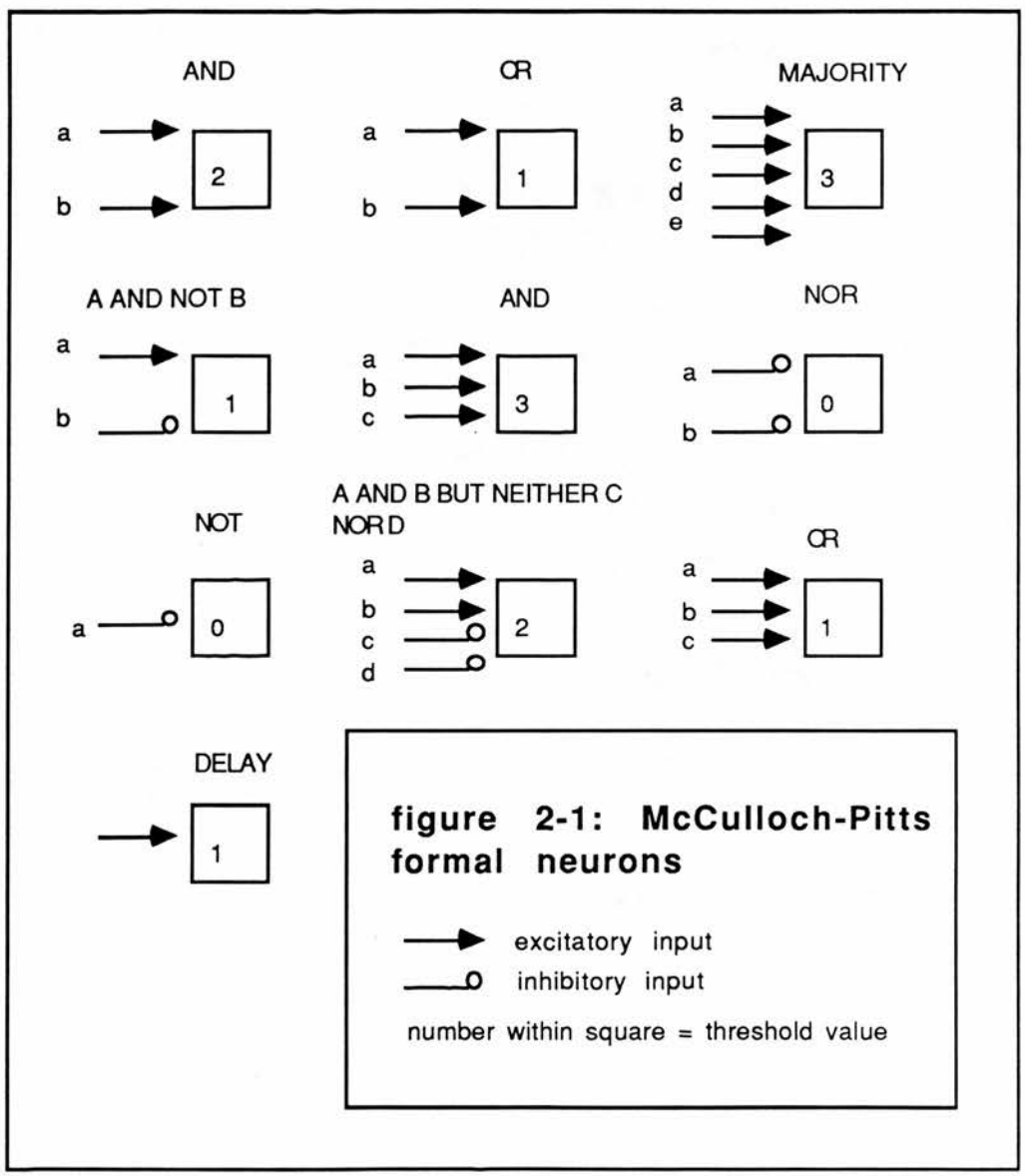


figure 2-1: McCulloch-Pitts formal neurons

12 McCulloch described a delay as a 'proposition on the move.' ". . . The time of occurrence of a signal (which is, if you will, no more than a proposition on the move). This was needed in order to construct theory enough to be able to state how a nervous system could do anything" (McCulloch, in Jeffress [1951, p. 36]).

Jerome Lettvin (1988, p. vii), who was a colleague of McCulloch at MIT in several research projects, recently described McCulloch and Pitts's (1943) paper as "the first attempt at a theory of how the mechanism of the brain can sustain mental process."[13] This motivation was acknowledged by Warren McCulloch himself (Jeffress, 1951, pp. 34-35) in the 1948 Hixon Symposium:

> "What we thought we were doing (and I think we succeeded fairly well) was treating the brain as a Turing machine . . . The delightful thing is that the very simplest set of appropriate assumptions is sufficient to show that a nervous system can compute any computable number."

McCulloch and Pitts' (1943) paper was of central importance in the development of the brain/machine issue in cybernetics. John von Neumann, a leading scientist of the post-war period who became well known for his contribution to the development of the stored-program computer, and a leading participant in the cybernetics movement, recognised the importance of McCulloch and Pitts' (1943) paper in the 1948 Hixon Symposium the following terms:

> "It has often been claimed that the activities and functions of the human nervous system are so complicated that no ordinary mechanism could possibly perform them . . . The McCulloch and Pitts result puts an end to this. It proves that anything that can be exhaustively and unambiguously described, anything that can be completely and unambiguously put into words, is ipso facto realizable by a suitable finite neural network. Since the converse statement is obvious, we can therefore say that there is no difference between the possibility of describing a real or imagined mode of behavior completely and unambiguously in words, and the possibility of realizing it by a finite neural network." (von Neumann, 1951, pp. 22-23)

---

[13] This point was also made by Papert (1965, pp. xxvi-xxvii): "They [McCulloch and Pitts, 1943] provide a definition of 'computing machine' that enables us to think of the brain as a 'machine' in a much more precise sense than we could before."

McCulloch and Pitts' (1943) paper was seen as showing that it was possible in principle to pursue the aim (central to cybernetics) of integrating the study of brain and machine. In practice, however, different approaches to that objective were developed, as I will show later in this section.

It is interesting to note that McCulloch-Pitts formal neural networks were used not only in neural network research, but also in theory of computation studies (Minsky, 1967) and in the development of the digital (von Neumann) computer. Furthermore, some researchers criticised attempts of developing McCulloch and Pitts' formal neural networks in the direction of neural network research. I will look at some of these developments briefly now.

Marvin Minsky (1967) used McCulloch and Pitts formal neural networks in his study of the theory of computation. Minsky studied the theory of finite-state machines (or finite automata, i.e. "machines which proceed in clearly separate 'discrete' steps from one to another of a finite number of configurations or states" [ibid., p. 11]) using McCulloch and Pitts formal neurons as his basic building elements.

As W. Aspray (1990, p. 173) pointed out, McCulloch and Pitts neural network terminology was also used, curiously enough, in von Neumann's first description of the stored-program computer (the 1945 EDVAC report). Turing's concept of a symbol manipulating machine was developed by John von Neumann and others in the design of the stored-program computer (so that the machine did not have to be reprogrammed for each new task).

Some researchers interpreted McCulloch and Pitts formal neural networks in an 'anti-neural network' direction. MIT AI researcher Seymour Papert, for example, in his introduction to (McCulloch, 1965) — a collection containing many of McCulloch's papers — criticised some researchers for using McCulloch and Pitts formal neural networks in 'attempts to dissolve the problem of

knowledge.' Papert's view is especially interesting because of his role in the history of neural network research (to be studied later). Neural netwoks is presumably one of the approaches criticised by Papert (1965, p. xxvii) for 'dissolving away the problem of knowledge:'

> "From this [McCulloch & Pitts, 1943] also follows the familiar flood of attempts to dissolve away the problem of knowledge into simple processes of cybernetic fiddling with thresholds or random program generators. Perhaps this is the place to emphasize that McCulloch is not to blame for this. Indeed, he insists that to understand such complex things as numbers we must know how to embody them in nets of simple neurons. But . . . we must . . . maintain a dialectical balance between evading the problem of knowledge by declaring that is 'nothing but' an affair of simple neurons, without postulating 'anything but' neurons in the brain."

In contrast with Papert's emphasis on (symbolic) knowledge, some cybernetics researchers spoke in favour of some kind of brain-style computing. In the 1948 Hixon Symposium John von Neumann made some comments in favour of a brain-like style of computing.[14] His point was that, although the 'complete and unambiguous' description in words of any phase of cognitive behaviour is in principle possible, for more complex cognitive behaviour such a description would be totally impractical (this problem is related to the AI problem of 'combinatorial explosion'). Von Neumann concludes his argument with a note in favour of some kind of brain-like computing.

> ". . . There is no difficulty in describing how an organism might be able to identify any two rectilinear triangles, which appear on the retina, as belonging to the same category 'triangle.' There is also no difficulty in adding to this, that numerous other objects, besides regularly drawn rectilinear triangles, will also be classified and identified as triangles — triangles whose sides are not fully drawn . . . etc. The more completely we attempt to describe

---

14 Von Neumann (1958) was very interested in the relationships between the computer/brain question.

> everything that may conceivably fall under this heading, the longer the description becomes. We may have a vague and uncomfortable feeling that a complete catalogue along such lines would not only be exceedingly long, but also unavoidably indefinite in its boundaries. Nevertheless, this may be a possible operation. All of this, however, constitutes only a small fragment of the more general problem of identification of analogous geometrical entities. This, in turn, is only a microscopic piece of the general problem of analogy. Nobody would attempt to describe and define within any practical amount of space the general concept of analogy which dominates our interpretation of vision . . . It is, therefore, not at all unlikely that it is futile to look for a precise logical concept, that is, for a precise logical description, of 'visual analogy.' It is possible that the connection pattern of the visual brain itself is the simplest logical expression or definition of this principle." (von Neumann, 1951, p. 24)

The tone of this remarks by von Neumann is not too far from the motivation of some neural network researchers of the 1950s and 1960s such as Frank Rosenblatt, as it will be shown in later sections.

The emphasis of cybernetics on information (as opposed to energy or matter), and the importance of the brain/machine issue, could be seen as a reaction to behaviourism, the dominant approach in psychology from the 1920s to the 1940s. Howard Gardner (1985, pp. 10-16) indicated this in his history of cognitive science, and stressed the 'symbolic' value of Karl Lashley's attack to behaviourism the 1948 Hixon Symposium on 'cerebral mechanisms in behaviour' (held in the California Institute of Technology, and published as [Jeffress, 1951]). Behaviourist psychologists had focused on observable (i.e. on stimulus/response level) behaviour, with an emphasis on conditioning and reinforcement. Questions of perception, representation, and cognition were far from the centre of their research agenda.

But it is important to note that the 'challenge to behaviourism' (using Gardner's term) from cybernetics was not as uniform as Gardner seems to indicate. Researchers started to study 'central' information-processing issues (i.e. what goes on in the cognitive system between stimulus and response), but different models of information-processing were proposed and analysed. Michael Arbib (1983, pp. 82-83) pointed out that five approaches to Wiener's above quoted programmatic goal of an integrated study of machine and nervous system emerged from the cybernetics movement. These were: (i) biological control theory (application of control theory techniques to the study of physiological systems, e.g. the control of the pupil); (ii) neural modelling (mathematical and microelectrode studies of single neurons or small nets of neurons, see Harmon & Lewis' [1966] comprehensive review of this area); (iii) AI (construction of computer programmes that exhibit aspects of intelligent behaviour, with an emphasis on symbol-processing), (iv) cognitive psychology (a central early contribution here is Miller, Galanter, & Pribram [1960]), (v) brain theory (a part of neural network research, what Lighthill [1973] later called computer-based central nervous system research). A more recent name for 'brain theory' is computational (or cognitive) neuroscience. Another approach was bionics, which studied the relationships between biological systems and engineering systems.

One important conclusion from this quick review of the cybernetic approaches to the brain/machine question is that symbolic AI and neural network research were only two approaches among many others. The variety of approaches to information-processing and the brain/machine problem within cybernetics can easily be shown by looking at the scientific conferences of the time. Examples of this are, to name but a few, the symposium on 'Mechanisation of Thought Processes' organised by the British National Physical Laboratory (NPL) in November 1958 (NPL, 1959) and the conferences on 'self-organisation' held in 1959 (Yovits & Cameron, 1960), 1960 (von

Foerster & Zopf, 1962), and 1962 (Yovits, Jacobi, & Goldstein, 1962). In the 'Mechanisation of Thought Processes' conference there were contributions from approaches including symbolic AI (M. Minsky and J. McCarthy), 'cybernetics' (D. M. MacKay, W. R. Ashby), pattern recognition (O. G. Selfridge, A. M. Uttley, W. S. McCulloch, and W. K. Taylor), and neural networks (F. Rosenblatt). In the 1962 conference on self-organisation there were contributions from perspectives including neural modelling (L. D. Harmon), brain theory/neural networks (W. S. McCulloch, M. A. Arbib, J. D. Cowan), neural networks (F. Rosenblatt, B. Widrow), neural networks/electrophysiological experiments (B. G. Farley), symbolic AI (A. Newell), and 'cybernetics' (D. M. MacKay). Within the cybernetics movement, work was done which was related to neural networks in diverse ways and degrees. Oliver Selfridge's (NPL, 1959, pp. 511-526) 'Pandemonium' system is an example; another one is work on pattern recognition in Britain by A. M. Uttley (NPL, 1959, pp. 119-147).

In the early and mid-1950s neural network research started to emerge from the cybernetics tradition. Two important elements of the work carried out by the first neural network researchers were McCulloch and Pitts' (1943) idea of networks of formal neurons and Canadian psychologist Donald Hebb's (1949) ideas of 'cell assembly' and learning by synaptic modification. In the rest of this section I discuss briefly some of these early attempts to combine McCulloch and Pitts neural networks and learning — the first neural network machines and simulations.

Donald Hebb published two important speculative hypotheses. One was that that the formation of networks of mutually excitatory neurons could be the basis of perception and representation processes.

> "It is proposed . . . that a repeated stimulation of specific receptors will lead slowly to the formation of an 'assembly' of association-area cells which can act briefly as a closed system after stimulation has ceased; this prolongs the time during which the structural changes of learning can occur and constitutes the simplest instance of

a representative process (image or idea)." (Hebb, 1949, p. 60)

With the notion of 'cell assembly,' Hebb hypothesised a link between perception (and representation) processes and their neural substratum. The equipotentiality of cell assemblies can be seen an antecedent of the notion of distributed, damage-resistant memory.[15]

The second speculative hypothesis by Hebb was about learning. Hebb suggested that learning happens by strengthening synaptic connections between neurons. This idea has been very influential throughout the history of neural network research. Hebb (ibid., p. 62) stated it in this way:

> "When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased."

Hebb's ideas of cell assembly and learning received considerable attention in the 1950s, and neural network researchers tried to test them using different approaches. One of these was digital computer simulation, which was possible after the advent of the digital computer in the early 1950s. B. G. Farley and W. A. Clark (1954) simulated a statistically connected network to investigate possible learning capabilities (Rosenblatt 1962a, p. 24). N. Rochester, J. H. Holland, L. H. Haibt and W. L. Duda (1956) carried out digital computer simulations of Hebb's ideas of the formation of cell assemblies and learning by synaptic modification.[16] It will be seen later that the development of the digital computer was very important in early AI and neural

---

[15] See this statement about equipotentiality: "The assembly is thought of as a system inherently involving some equipotentiality, in the presence of alternate pathways each having the same function, so that brain damage might remove some pathways without preventing the system from functioning, particularly if the system has been long established, with well developed synaptic knobs which decrease the number of fibers that must be active at once to traverse a synapse" (Hebb, 1949, p. 74).

[16] Rochester was manager of information research at IBM's research centre in Poughkeepsie, New York (McCorduck 1979, p. 93).

networks. Nevertheless, it is important to emphasise that the time of early neural network research was only the beginning of the development of digital computer technology. In later sections it will be seen that implementation (i.e. the construction of neural network machines) was a very important element of the main early neural computing developments in the late 1950s and early 1960s.

One neural network project carried out in the 1950s is especially interesting for this dissertation. It is a machine built by Marvin Minsky and Dean Edmonds in 1951 at Harvard University.[17] This machine was one of the first neural network machines ever built, perhaps the first one. Minsky's work on neural networks is of particular interest here because Minsky later became one of the most influential critics of neural networks. Minsky described that project to me in the following terms:

> "I was interested in the problem of making a learning machine by modifying the connections of the network. I built a machine in the summer of 1951. It was not perceptron-like. I guess you could call it a machine with Hebb's synapses. It was a random network without layers. When you put a signal into it, some inputs, they spread through the network probabilistically. The threshold of each neuron was the probability that the neuron would fire. Then, after it did something, you decided whether it was good. If you rewarded it, it increased the connections that were used, and, if you punished it, it lowered the coefficients. But it only affected the neurons which actually conducted, the synapses that transmitted. I was trying to do things like recognising something in different positions, but it couldn't, because it would take too many neurons. It learned some things and didn't learn others. I realised that it was too much work to do." (Minsky, interview)

Another account by Minsky about his random neural network machine can be found in in an interview carried out by Jeremy Bernstein (1981, pp. 69-70) in 'The New Yorker:'

---

[17] Edmonds was a young graduate student in physics at the time (Bernstein, 1981, p. 69).

". . .'Somehow [says Minsky], he [George Miller from Harvard University] managed to get a couple of thousand dollars from the Office of Naval Research, and in the summer of 1951 Dean Edmonds and I went up to Harvard and built our machine. It had three hundred [vacuum] tubes and a lot of motors. It needed some automatic electric clutches, which we machined ourselves. The memory of the machine was stored in its control knobs — forty of them — and when the machine was learning it used the clutches to adjust its own knobs.' Many of the networks were wired at random, so that it was impossible to predict . . . 'Because of the random wiring [says Minsky], it had a sort of fail-safe characteristic. If one of the neurons wasn't working, it wouldn't make much of a difference — and, with nearly three hundred tubes and the thousands of connections we had soldered, there would usually be something wrong somewhere. In those days, even a radio set with twenty tubes tended to fail a lot. I don't think we ever debugged our machine completely, but that didn't matter. By having this crazy random design, it was almost sure to work, no matter how you built it.' . . ."

Minsky continued to work in neural networks for a few years.[18] However, soon after he finished his PhD, Minsky started to change his research interests towards symbolic artificial intelligence. He thought that further progress in neural networks would be too difficult.

"One reason why people didn't study neural nets is because it was too hard. But also around this time, 1950, I began to think about the high level processes, like how do you solve a problem, how do you do reasoning, how do you make a machine that solves a geometry problem. I started to think about symbolic artificial intelligence, how do you make a machine to do that. I worked on that from 1956." (Minsky, interview)

". . . 'Finally I decided that either this [neural networks] was a bad idea or it would take thousands of millions of neurons to make it work, and I couldn't afford to try to

---

18 Minsky's (1954) PhD dissertation was on 'Neural nets and the brain-model problem.'

build a machine like that.' . . ." (Minsky, in Bernstein, 1981, p. 70)

In 1956 (the year Minsky refers to above) there were a series of meetings, workshops, and informal contacts at Dartmouth College (Hanover, New Hampshire). These meetings are taken as the starting point of symbolic artificial intelligence as a distinct scientific specialty. It is interesting to note that from the four researchers who applied to the Rockefeller Foundation for financial support for the Darmouth meetings — namely John McCarthy, Marvin Minsky, Nathaniel Rochester and Claude Shannon (McCorduck, 1979, p. 93) — there were two — Minsky and Rochester — who had been doing research on neural networks previously. Shannon had made important contributions to information theory in the 1940s, and McCarthy (at the time lecturer of mathematics at Darmouth) soon became a leading symbolic AI researcher (in 1960 he formulated the LISP programming language). Other researchers who at some point attended the Darmouth conference were Arthur Samuel, Oliver Selfridge, Allen Newell, and Herbert Simon.

It was around this time, in the early and mid-1950s, that Minsky decided to work within the symbol-processing approach to AI, and abandoned the idea of building intelligent machines using neural networks. It cannot be said, however, that he completely abandoned neural network research. It will be seen in later sections that he became one of the most influential critics of neural network research (Minsky & Papert, 1969). I will come later (section 2.5) to the issue of the emergence of symbolic AI, and the role which symbol-processing AI researchers played in the evolution of early neural network research. For now it is important to bear in mind that the emergence of symbolic AI started to gain momentum in the late 1950s and early 1960s, just at the same time as the main early neural network projects (to be studied in the coming sections) were being developed. So even though Minsky and other important cybernetics researchers saw no promise in neural network research in the mid-1950s and started to work firmly within the symbol-processing framework,

other researchers interested in the brain/machine problem did not share their view, and pursued the neural network approach.

In this section I have looked at the origins of neural network research in the cybernetics movement of the 1940s and 1950s. I have shown that the brain/machine problem, a central one in cybernetics, was developed in different directions. I have also shown that McCulloch and Pitts' (1943) paper, which is usually taken as the beginning of the history of neural networks, was of central importance in brain/machine cybernetics research. McCulloch and Pitts formal neurons were used in a variety of contexts, and not only in neural network research. Towards the early and mid-1950s, researchers started to experiment with neural networks with modifiable connections (inspired by Hebb, as well as McCulloch and Pitts). This can be seen as the beginning of neural network research. I have concluded by pointing out that, towards the mid- and late 1950s symbol-processing AI and neural network research were emerging from the cybernetics tradition.

2.2 The computer and the brain

In this section I look at some of Frank Rosenblatt's ideas about the brain/machine problem in a general and introductory way. Rosenblatt was one of the most important early neural network researchers. Among the early neural network researchers of the late 1950s and 1960s, he was the one who elaborated the general ideas guiding the 'neural network enterprise' in their most explicit way. In sections 2.3 and 3.2 I study Rosenblatt's work on his perceptron machine in more detail. In chapter three I show that Rosenblatt was also the 'symbolic leader' of the neural network position in the perceptron controversy, an event of great importance in the evolution of early neural network research.

Rosenblatt's ideas on the brain/machine theme are studied in this section within a 'metaphor scheme.' First I discuss the sense in which I use the metaphor scheme here. Afterwards, I study how Rosenblatt used the brain as a metaphor for studying cognition and building artificial intelligence systems. Rosenblatt's 'neural inspiration' is analysed by looking both at the localisation/distribution issue and at the randomness issue. Finally, Rosenblatt's ideas about the von Neumann computer are examined. In this section I look at some of the the ideas behind early neural network research in a general and introductory way. It is interesting to note that many of the ideas which became popular after the re-emergence of neural network research in the late 1980s had already been explicitly formulated by Rosenblatt in the late 1950s and early 1960s.

The 'metaphor scheme' has been used as interpretative device in studies of science in two senses. On the one hand, elaborating on T. Kuhn's (1970) notion of exemplar and M. Hesse's (1963) work on analogies in science, B. Barnes (1974) developed a model for interpreting scientific activity based in the use (by scientists)

of accepted problems/solutions as metaphorical resources in redescribing and solving new puzzles. I will use this notion in the analysis of some innovations in neural networks in the 1980s (section 4.2: 'Networks with symmetric connections: metaphors and innovation in neural computing'). The second sense in which the metaphor scheme has been used in studies of science is a more general one. Barnes (1974, pp. 97 and 93) was using the metaphor scheme in this sense when he talked about "fundamental changes in the way of acting and perceiving taken for granted within a specialty, associated with replacement or major alteration in the dominant model or metaphor", or about a "major reorganization of the general model or metaphor at the base of the activities of a specialty."[19] In this section I will use the term 'metaphor' in this second, more general sense (but I will not talk about replacement of metaphors or reorganization of research areas).

This second sense of 'metaphor' was developed by Donald Schon and others. Schon (1963) used a metaphor scheme in his studies of the the evolution of ideas. Schon suggested that 'concept displacement' (that is the use of concepts as metaphorical resources) is one of the main sources of change and innovation in the production of ideas. When a metaphor is used, an old concept or set of concepts is taken from one context and used as a resource in a new context or situation. In the process of being used as an interpretative resource in a new situation, the meaning of the old concept itself changes (evolves).

> "New concepts do not spring from nothing or from mysterious external sources. They come from old ones . . . [This] is possible if new concepts emerge out of the

---

[19] Barnes (1974, pp. 146-147) pointed out that the metaphorical character of scientific knowledge (as of knowledge and thought in general) shows its culture-bound character (and therefore the possibility of studying its production and development from a sociological point of view): "There are two ways in which the initial selection or construction of a model and its institutionalization, may be related to the social context. First, the stock of available cultural resources is a function of the milieu and the range of actors' experience within it. Second, what counts as a part of accepted knowledge or an accepted standard of judgement will again depend on the milieu, and may depend on actors' social roles, and the concerns and interests of the groups to which they belong."

> interaction of old concepts and new situations, where the old concept is not simply re-applied unchanged to a new instance but is that *in terms* of which the new instance is seen. This is what we have described as the displacement of concepts — a process in which old concepts, in order to function as projective models for new situations, come themselves to be seen in different ways . . . The culture provides us with the informal theories from which our formal theories are displaced" (Schon, 1963, pp. 192 and 196)[20]

In section 2.1 it was seen that neural network research (and neuroscience-oriented neural networks) was one of the approaches which emerged from the cybernetic concern with the relationships between brain and machine. Others included neural modelling, biological control theory, and symbolic AI. Two of these approaches were especially concerned with the problem of artificial intelligence (i.e. the problem of building machines that exhibit cognitive and intelligent abilities), namely symbolic AI and neural networks.

Symbolic AI and neural networks were based on different general models or 'metaphors' (in the second sense mentioned above). For symbolic AI the digital computer was a metaphor of cognitive processes, that is cognition was understood as computation, and computation was defined as symbol manipulation. Emphasis was given to the transformation of symbolic expressions in ways sensitive to the logico-syntactic structure of those expressions (as in inference and formal reasoning). In symbolic AI, the level of symbol processing (i.e. the level of cognition) was clearly separated from the level of (neural or hardware) implementation. Recently P. Johnson-Laird (1988, pp. 7-8) described the metaphor at the basis of symbol-processing AI and cognitive science as follows:

> ". . . The invention of the programmable digital computer, and more importantly its precursor, the mathematical theory of computability, have forced people to think in a

---

[20] Schon (1963) also made some interesting comments on the use of metaphors in explaining mind throughout the history of Western civilisation.

> new way about the mind. Before computation, there was a sharp distinction between brain and mind: one was a physical organ, the other a ghostly nonentity that was hardly a respectable topic of investigation . . . After computers, there can be no such scepticism: a machine can be controlled by a 'program' of symbolic instructions, and there is nothing ghostly about a computer program. Perhaps the mind stands to the brain in much the same way that the program stands to the computer. There can be a science of the mind."

The advent of the digital computer in the 1950s and its development in the 1960s favoured the symbol-processing approach to AI. But the 'computer metaphor' was not the only model used by AI researchers in the 1950s and 1960s. The brain/machine issue of cybernetics was developed in another direction too. Early neural network researchers went the opposite way. Instead of using the computer as a metaphor of cognition, they used the brain itself as a metaphor in studying cognitive processes and in designing and building intelligent computers.

It was said above that, when using a (cultural, conceptual or technological) resource as an interpretative resource in a new situation, the resource being so used is itself transformed. Something like this can be said of Rosenbaltt's use of knowledge from neurophysiology and neurobiology for designing his computing systems (or 'perceptrons'). In using the brain as a metaphor, the brain was transformed into a simplified model of the brain. Thus using the brain as a metaphor does not mean that the explanation and modelling of intelligent and cognitive processes belongs to the the neurophysiological or neurobiological level, or that computational models of *actual* neural structures or processes are to be built. Rosenblatt was *inspired* by contemporary brain studies. He used contemporary brain studies as an inspiration to design and build artificial (computational) models of intelligence and cognition, not to model the brain itself. Concepts from the brain sciences were

used in a different context of scientific activity, namely (what today is called) AI and cognitive science.

Rosenblatt was aware that his 'brain models' or 'perceptrons' (these two terms were equivalent for him) were based on 'extreme simplifications' of real nervous systems. These extreme simplifications seemed to him necessary in order to fill the gap between brain and cognitive processes.

> "Perceptrons are not intended to serve as detailed copies of any actual nervous system. They are simplified networks, designed to permit the study of lawful relationships between the organization of a nerve net, the organization of its environment, and the 'psychological' performances of which the network is capable . . . More likely, they represent *extreme simplifications* of the central nervous system, in which some properties are exaggerated, others suppressed." (Rosenblatt, 1962a, p. 28, emphasis added)

Rosenblatt acknowledged that the elements of his machines, i.e. McCulloch and Pitts formal neurons with modifiable connections, were also a simplification of actual brain cells.

> "Simplifications will therefore be introduced, as in the manner of the McCulloch-Pitts neuron; but it should be remembered that the biological neuron is considerably more complicated, and may incorporate within itself functions which we require whole networks of simplified neurons to realize." (ibid., p. 35)

Knowledge from the brain sciences was used in early neural networks in a 'metaphorical' way. A quick look at some of the other specialties dealing with the brain/machine problem which emerged from cybernetics helps confirm this. Early neural network researchers such as Frank Rosenblatt were closer to AI research than to specialties which sought to model actual brain mechanisms and processes, such as 'neural modelling.' The difference between neural network research and the modelling of actual neural mechanisms and processes was stressed in the mid-1960s by L. D. Harmon and E. R. Lewis (1966) in their review

(published in 'Physiological Reviews') of neural modelling research. Harmon and Lewis complained that he term 'neural modelling' was sometimes being used to refer both to neural network research and to research aiming at modelling actual biological or physiological structures and processes. Harmon and Lewis' emphatic distinction between the two research areas (neural networks and neural modelling) confirms the (above mentioned) 'displacement' typical of the use of metaphorical resources in developing new scientific ideas.

> "The formal neuron is presently being deprecated as being an extremely unrealistic simplification of a biological neuron and thus insufficient as a model . . . There are two radically distinct research areas to which the term 'neural modelling' has been applied. In one the intent is to represent the physiological phenomena . . . In the other area [neural networks], often euphemistically called neural modelling, the network properties of systems of quasi-neural elements are explored. The intent is to build automata whether or not they replicate in realistic detail any actual physiological functions. Such is the province of 'adaptive systems' or 'self-organizing systems' [i.e. neural networks]. In most cases only a few selected neural properties are adopted simply to see what can be done by applying mathematical or computer concepts to neuron-like elements . . . Most of these systems employ elements that are simple, time stationary, formal neurons, that is, threshold elements that lack the many temporal dependencies of biological neurons. Terms like 'neuron' and 'synapse' are used loosely and irrelevantly. Further, in clear distinction to real nerve nets, these systems start with completely chaotic (random) connection patterns. Finally, the mechanisms for network change are based on Hebb's still unproven postulate of synaptic change for memory and functional modification . . . The fundamental problems being addressed by self-organizing systems research are certainly interesting, but appear to be neurophysiologically irrelevant." (Harmon & Lewis, 1966, pp. 515, 517, and 575-576)

Researchers within the neural modelling approach concentrated on the modelling of single neurons or small networks of neurons

using mathematical, electronic, and analog and digital computer techniques (ibid.). The neural network approach was rather different, because neural network researchers were from the beginning concerned with the problem of perception and cognition (and therefore representation, although of a non-symbolic kind). They studied the behaviour of large networks of interacting, simplified neuron-like elements, and their aim was to show how certain cognitive or intelligent behaviour could emerge from such a physical system.

There are striking similarities between Rosenblatt and neural network researchers of the 1980s with respect to the notion of neural inspiration. A good part of current neural network research is very much within the 'brain as metaphor' tradition.

> "Thus, we have, by and large, not focused on *neural modelling* (i.e., the modelling of neurons), but rather we have focused on *neurally inspired* modelling of cognitive processes." (Rumelhart & J. McClelland, 1986a, p. 130)[21]

A brief discussion of the details of Rosenblatt's neural inspiration is necessary in order to understand both Rosenblatt's theoretical and methodological assumptions and the architecture of the 'neural computer' that his group built at Cornell Aeronautical Laboratory. Some of the main characteristics of Rosenblatt's neural inspiration can be discussed around the issue of distribution/localisation of brain function. Rosenblatt openly admitted that he was impressed by the work of Harvard University neuropsychologist Karl Lashley on distribution of brain function. Nevertheless, he was aware that localisationist hypotheses were gaining strength in contemporary

---

[21] P. Smolensky's (1988) recent move to call neural network research the 'subsymbolic' paradigm could also be understood in this context. He claimed that the neural network approach studies and models intelligence at the 'subsymbolic,' that is at a level somewhere between the brain level and the symbol-processing level. But although at the time of the recent resurgence of neural networks in the second half of the 1980s this was the main view of the level at which neural networks should study cognition and build intelligent machines, other important researchers related to neural networks (including Carver Mead and Christoph Koch) were (and are) attempting to mimic neurobiology as closely as possible.

neurophysiological studies, mainly in 'low level' (or 'sensory', in Rosenblatt's terms) tasks such as vision. Let me look at the distribution/localisation issue in more detail now.

Gardner (1985, pp. 268-271) pointed out that by the late 1940s (the time of cybernetics) the debate about localisation of brain function — which had been closed in favour of localisationism in the early years of the 20th century (Star, 1989a) — had (to some extent at least) reopened again, and that there was a resurgence of 'holistic' positions in some research circles in neuroscience and psychology (Gardner mentions neurologists like P. Marie, K. Goldstein, H. Head, and the influence of Gestalt psychology, as well as the influence of Lashley himself). Nonetheless, Star (1989a, ch. 7) indicated that localisationism has been dominant in many other brain research specialties (including neurophysiology, neurosurgery, psychiatry, and psychosurgery) over the years after the closure in favour of localisationism in the first decade of the 20th century, and that diffusionist work such as Lashley's was rather isolated (ibid., p. 28).

But even if Lashley's work on the 'equipotentiality and plasticity' of brain function was far from dominant in the neurosciences as a whole, it was important for early neural network research because it emphasised the distributed nature of memory.[22] The following comment by Minsky and Papert (1969, p. 19), who criticised the use of distributed memory systems in AI-like research, confirms this:

> "It was a great disappointment, in the first half of the twentieth century, that experiments did not support nineteenth century concepts of the localization of memories (or most other 'faculties') in highly local brain areas . . . [Those experiments] . . . lead to a search for nonlocal machine-function concepts."

---

[22] This influence might have been increased by Lashley's role in the brain/machine discussions within cybernetics (in which researchers like von Neumann and McCulloch often participated) and his criticism of behaviourism (a few comments on this were made in the previous section). See the 1948 Hixon Symposium at Caltech (Jeffress, 1951). Among the participants of the Hixon Symposium were Lashley, Gestalt psychologist Wolfgang Köhler, von Neumann, and McCulloch.

After carrying out brain lesion experiments, Lashley (1950, pp. 62-63) came to the conclusion that, within the regions of the brain, it was sometimes more important how much tissue was removed than where the tissue was removed from (i.e. its topological location):

> "It is not possible to demonstrate the isolated localization of a memory trace anywhere within the nervous system. Limited regions may be essential for learning or retention of a particular activity, but within such regions the parts are functionally equivalent. The engram is represented throughout the region . . . Every instance of recall requires the activity of literally millions of neurons. The same neurons which retain the memory traces of one experience must also participate in countless other activities."

Lashley's hypothesis of equipotentiality favoured early neural network researchers' notion of distributed memory, as can be seen in comments by Rosenblatt himself and his colleague David Block.

> "At the time that the first perceptron model was proposed, the writer was primarily concerned with the problem of memory storage in biological systems, and particularly with finding a mechanism which would account for the 'distributed memory' and 'equipotentiality' phenomena found by Lashley and others." (Rosenblatt, 1962a, p. 4)

> "There is a certain equi-potentiality involved in brain functions (Lashley, 1929) . . . It seems clear that memory and the other higher functions are distributed in the fine structure of the brain." (Block, 1962, p. 139)

Rosenblatt (1962a, pp. 41–42) recognised the importance of localisation in sensory processes, but emphasised the distributed nature of higher cognitive functions such as 'thinking' and 'memory.'

> "The extreme hypothesis of equipotentiality advocated originally by Lashley (1929), (who observed that cortical ablation appeared to produce a general deficit in performance proportional to the amount of cortex

> extirpated, rather than eliminating specific memories and abilities) has been modified in the direction of relative localization, which is quite strict for certain sensory functions, and comparatively weak and readily modified for more complicated control functions, thinking and memory."

More aspects of the distribution issue will be discussed later. Now I would like to concentrate on the randomness issue. One distinctive feature of Rosenblatt's perceptron was the random character of some of its connections (this will be studied in section 2.3). Rosenblatt believed that the fine structure of the brain's neurons and interconnections was random.

> "While recent work on localization has shown some surprisingly precise mapping of functions, modern morphological investigations have borne out the apparently statistical organization of the 'fine structure' of neurons and their interconnections." (Rosenblatt, 1962a, p. 21)

In the randomness issue Rosenblatt was motivated by both neurophysiological and methodological considerations. Some brain researchers of the 1950s had emphasised the statistical, random character of the fine structure of the nervous system. One example of the type of brain research which inspired Rosenblatt and colleagues in their notion of the statistical and random character of the circuitry of the brain is D. A. Sholl's work in the 1950s. Sholl (1956, pp. 76 and 111) related the statistical character of the neural connections in the brain to the distributed character of brain function:

> "There are small portions of cortex mainly associated with the arrival of impulses (the sensory side), and there is a portion of cortex mainly concerned in the initiation of impulses leading to changes in the musculature; there remains the large mass of cortex to which no special 'function' can be ascribed and, so far as is known, no distinctive histological organization . . . Although the number of connections made by an individual neuron cannot be stated explicitly, the connective fields of the neurons may be defined and measured in statistical terms. Such a description corresponds to the plasticity of response shown in physiological and psychological experiments and

> which is unaccountable in terms of unique circuits. In the past, attempts have been made to correlate certain psychological concepts such as 'memory' and 'intelligence' with the activity of circumscribed regions of the cortex. When these concepts are defined operationally in a manner that enables them to be measured, experimental work has shown little evidence for regional specialization of this kind and one is driven to the opinion that these properties are general attributes of all cortical tissue. Such a conclusion is in accordance with the anatomical evidence."

It is important to emphasise that Rosenblatt thought that by postulating random connections in his machine he was not contradicting neuroscience research.

> "Specific connections cannot be traced with sufficient precision in nervous tissue to say whether or not a particular wiring diagram is exactly realized." (Rosenblatt, 1962a, p. 18)

Rosenblatt and Block believed that, given the huge numbers of cells and interconnections between cells in the brain, its circuitry could not be specified in all its details genetically. Postulating randomness in the the neural connections seemed to them a feasible solution that did not violate contemporary anatomical evidence (remember that we are talking about the mid-1950s here). The issues of randomness and distribution were linked to what today is called 'graceful degradation.' In neural networks, unlike in digital computers, functioning does not degrade completely if a few connections do not work properly.

> "To establish detailed anatomical information regarding the connections of the 10^{10} neurons in the brain presents a formidable laboratory task . . . It seems impossible to map the entire topology of the [brain's] neural network. Moreover, even if we accomplished this, we would then face the disheartening task of analyzing the performance of such a network. Now in a digital computer every connection must be exact or the answer can be entirely wrong . . . However, it is clearly not true that the connections must be exactly right for the brain to function at all

> . . . Furthermore it seems unlikely that the genes would carry the information to specify every one of the 10^{13} connections. It seems more plausible that only certain parameters of growth are specified and the fine connections are grown in a more or less *random manner*, subject to these constraints. Thus the detailed connection scheme would be unique to each individual. If it is true that individuals, with connection schemes specified only by certain parameters of growth, function in similar ways, then there is hope that the performance of such a system might be analyzed in terms of such parameters. This also implies that the operation of the brain is radically different in principle from the logical circuitry of digital computers." (Block, 1962, pp. 140-141, emphasis added)

Rosenblatt and colleagues thought that, by using random connections, they were making a safe methodological assumption, because they did not commit themselves to a particular pattern of connections. In other words, unlike the von Neumann computer, their machine's performance did not depend on specific details of its wiring circuits.

> ". . . There is no particular virtue in randomization. This is precisely the reason it is used! If a randomized net can learn, then certainly so can a net with carefully specified connections . . . So in the sense of a 'worst case' or 'minimal constraint' feasibility study, one randomizes the connections." (Block, 1970, pp. 514-515)

Rosenblatt's neural inspiration can be summarised by looking at the similarities between two of his graphics. One (figure 2-2 below) represents his view of the basic topological structure of the nervous system. The other one (figure 2-3 below) is a representation of his 'general experimental system.' The machine that Rosenblatt's group actually built (to be studied in section 2.3) appears in bold in figure 2-3.

Figure 2-2 (after Rosenblatt, 1962a, p. 37) shows that in Rosenblatt's simplified topological structure of the nervous system there is a combination of localisation of function in the sensory and motor parts (represented in the figure by smaller

circles) and distribution in the central or association network (represented in the figure by a bigger circle). It was said earlier that Rosenblatt emphasised the importance of the distributed, association areas. In his view, they were the physical support of higher functions such as memory or speech. The following remark by Rosenblatt (1962a, pp. 40-41) summarises his view of the localisation/distribution issue, which is reflected in graphics 2-2 and 2-3:

> "The functional organization . . . has been most firmly established in the case of sensory and motor tracts, where a particular position in the brain is correlated with a particular sensory locus, or a particular set of muscles whose activity it controls . . . One feature which is of particular importance . . . is the apparent plasticity of localization in the 'association areas' in contrast to the relatively fixed and irreplaceable character of the sensory and motor tracts."

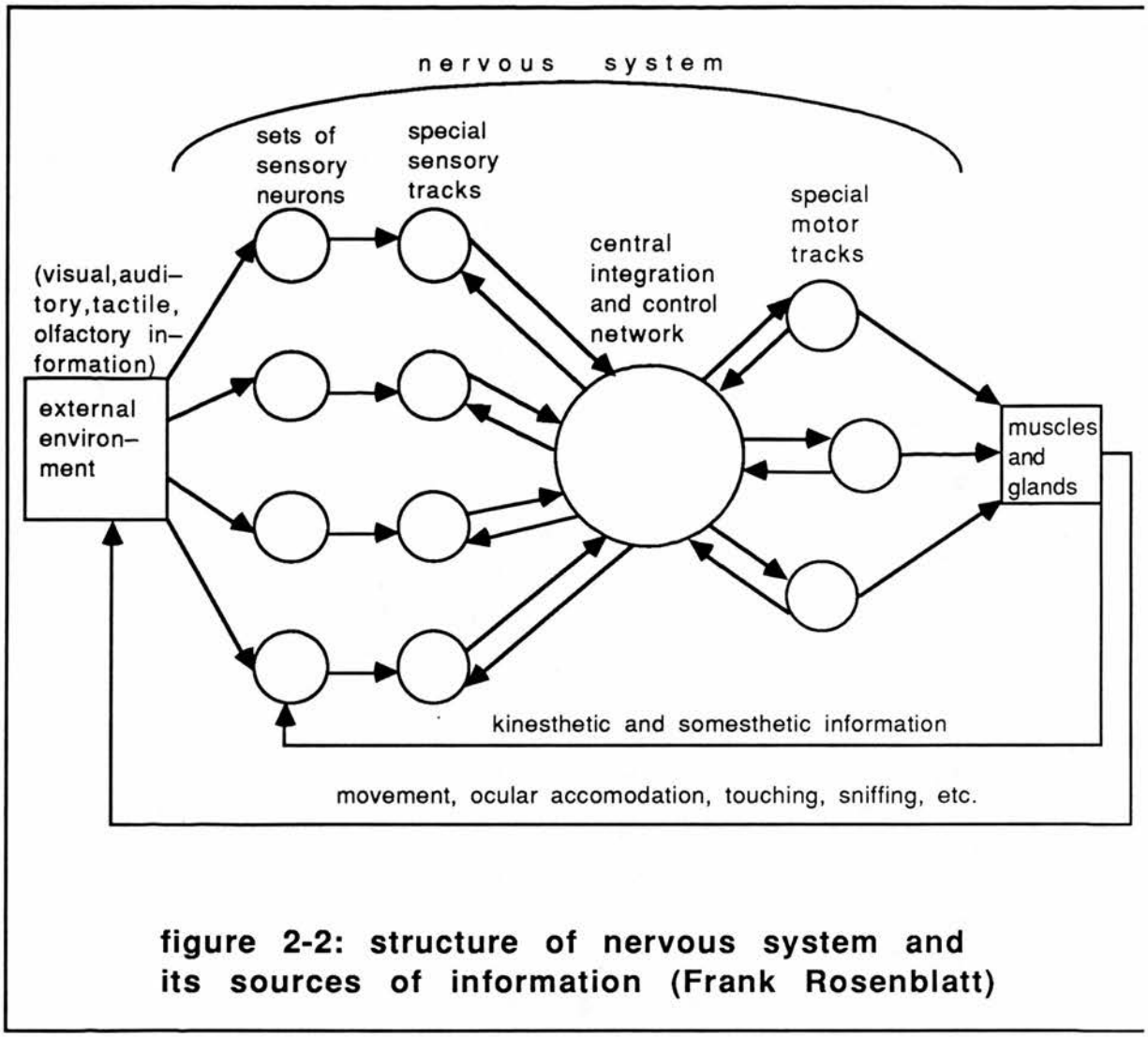


figure 2-2: structure of nervous system and its sources of information (Frank Rosenblatt)

Figure 2-3 (after [Rosenblatt, 1962a, p. 63], and [Rosenblatt, 1962b, p. 388]) shows the general characteristics of Rosenblatt's neural network machine design. There are three layers of units: sensory, association, and response. The arrows in both figure 2-2 and figure 2-3 represent direction of connections, and not actual connections. In the perceptron (in bold in figure 2-3) all the connections were feedforward. The connections between units in the perceptron are not represented in figure 2-3. One difference between the nervous system and the experimental system is that the latter needs a reinforcement control system (i.e. a supervisor, see figure 2-3) in order to improve its performance (learning). I will describe the perceptron in detail in section 2.3.

Figures 2-2 and 2-3 show, at a general level, the similarity between Rosenblatt's view of the nervous system and his general design for neural network machines (what this means in detail will be seen in later sections). Rosenblatt's use of the brain as a metaphor in the construction of information-processing machines becomes apparent by comparing those two figures.

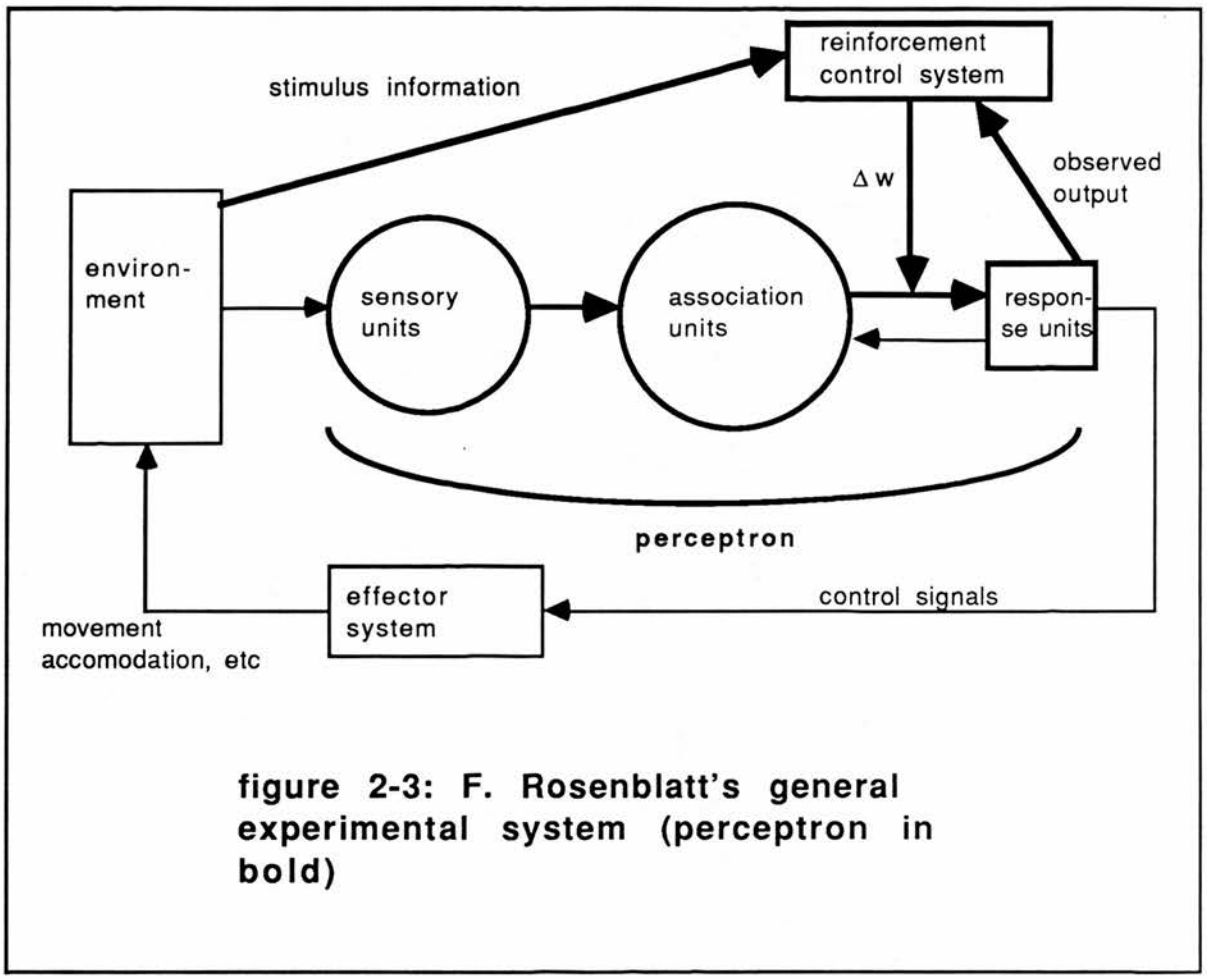


figure 2-3: F. Rosenblatt's general experimental system (perceptron in bold)

It has been seen so far that Rosenblatt used the brain as a metaphor for building neural network machines. It was said earlier that symbol-processing AI researchers had taken the opposite approach: they were using the von Neumann computer as a metaphor for studying cognition and building intelligent machines. Rossenblatt opposed the 'von Neumann computer metaphor' explicitly and actively. He made frequent and explicit

comments on this both inside and outside the research community. The following comment appeared in 'The New Yorker' magazine.

> "Digital computers, Dr. Rosenblatt said, are equipped to solve certain problems more quickly and accurately than human beings can, but the problems must be prepared and, in effect, spoonfed to them by specialists; being basically adding machines, the computers lack creativity. The distinctive characteristic of the perceptron is that it interacts with its environment, forming concepts that have not been made ready for it by a human agent . . . Both computers and perceptrons have so-called memories; in the latter, however, the memory isn't a mere storehouse of deliberately selected and accumulated facts but a free, indeterminate area of association units, connecting, as nearly as possible at random, a sensory input, or eye, with a very large number of response units." (The New Yorker, 1958, p. 45)

The issue of distributed memory, with each processing unit participating in many representations and each representation being formed by the activation of many units, was at the basis of Rosenblatt's opposition to the von Neumann computer (in a von Neumann computer data are stored in discrete, unrelated locations).[23] Rosenblatt's opposition to the von Neumann computer was of course related to his opposition to symbol-processing AI.

> "Theorists are divided on the question of how closely the brain's methods of storage, recall, and data processing resemble those practised in engineering today. On the one hand, there is the view that the brain operates by built-in algorithmic methods analogous to those employed in digital computers, while on the other hand, there is the view that the brain operates by non-algorithmic methods, bearing little resemblance to the familiar rules of logic and

---

23 ". . . The view that the brain contains its memories in a widely dispersed, intermingled form, suggests a mechanism in which the same cells participate in a great variety of different, and perhaps totally unrelated, memory organizations. A model which can separate distinct memories from such a multiply overwritten system will be quite different in character from one in which each remembered event is stored in its own distinct location." (Rosenblatt, 1962a, p. 59)

> mathematics which are built into digital devices" (Rosenblatt, 1962a, p. 10)

Among the properties of distributed memories is 'graceful degradation' (using today's terms). Rosenblatt emphasised the differences in this respect between the brain and the digital computer, where 'in some cases a single misconnection would be sufficient to make the system inoperable' (Rosenblatt, 1962a, p. 17).

> "The models which conceive of the brain as a strictly digital, Boolean algebra device, always involve either an impossibly large number of discrete elements, or else a precision of the 'wiring diagram' and synchronization of the system which is quite unlike the conditions observed in a biological nervous system." (Rosenblatt, 1959, p. 422)

The methodological differences between the symbol-processing approach and the neural network approaches were also stressed by Rosenblatt. He argued that, whereas in the symbol-processing approach one needs an accurate description of the behaviour to be modelled before computer models can be actually built, in a neural network system cognitive behaviour emerges from the physical system itself.

> "In the monotypic [i.e. symbol-processing] approach, the theorist generally begins by defining as accurately as possible the performance required from his model . . . Given a description of the required performance the theorist then proceeds to design a computing machine or control system embodying the required function . . . In the genotypic [i.e. neural network] approach . . . the organization of the network is specified only in part, by constraints and probability distributions which generate a class of systems rather than a specific design . . . In the monotypic approach, the functional properties are generally postulated as a starting point. In the genotypic approach, they are the end-objective of analysis, and the physical system itself (or the statistical properties of the class of systems) constitutes the starting point. This means that psychological functions need not be fully determined in full

> detail before setting out to construct a model." (Rosenblatt, 1962a, pp. 11-12 and 19-20)

In the quotation above Rosenblatt insists again on the 'loose' organisation of his system as compared to the organisation of a von Neumann computer. In other statements, he even talked about 'freedom of connections.'

> "A perceptron . . . is usually characterized by the great freedom which is allowed in establishing its connections, and the reliance which is placed upon acquired biases, rather than built-in logical algorithms, as determinants of its behavior." (Rosenblatt, 1962a, p. 5)

This is related to the issue of randomness discussed earlier. Block (1970, p. 515) indicated that the origin of the idea of randomness in perceptron systems goes back to cybernetics (to researchers like K. Craik, W. S. McCulloch and W. R. Ashby). The idea of randomness could perhaps be related to what V. Pratt (1987) called the 'political background' of cybernetics.

> ". . . Cybernetics began in effect as a joint research project between physiology and control engineering. But there was more colour to it than that implies. It was launched and conducted within an explicit ideological framework . . . It was a political outlook that belonged very much to the 1940s . . . Politically, those ranged against Hitler defined themselves as essentially opposed to 'dictatorship,' and the founders of cybernetics saw their ideas as developing a new, non-dictatorial conception of 'control'. In societies as in smaller organizations, their conception presented an alternative to the assumption that efficiency depended on a central intelligence responsible for decision making, and an army of agents ready only to carry the decisions through . . . For stability, homeostasis, there needed to be speedy and plentiful movement of information throughout a system's components, and this is just what failed in the mass market society, where the ostensible means of spreading information — books, newspapers, radio — were corrupted by the profit imperative." (Pratt, 1987, pp. 200-201)

This 'political background' seems to harmonise better with a distributed, neural network-like model of information-

processing rather than with a serial, von Neumann one.[24] This 'political metaphor' theme could perhaps be studied further in social studies of cybernetics. Some work of this kind is being done in areas such as distributed AI.[25] Nevertheless, I did not consider practical to follow a 'political metaphor' kind of hypothesis here for two reasons. First because more sociological and historical studies of cybernetics — and of the relationships between cybernetics and neural networks — are needed before such a hypothesis can be studied seriously. And secondly because the more cybernetics-flavoured elements of Rosenblatt's work (such as the random character of some of the connections of the perceptron) were not taken up by other early neural network researchers.

In this section I have looked at some of Rosenblatt's general ideas in a general and introductory way. I have used a 'metaphor scheme' to examine certain aspects of Rosenblatt's neural 'inspiration' (or 'brain-style' information processing). I have also discussed Rosenblatt's opposition to the von Neumann computer.

---

[24] More recently other neural network researchers used terms like 'collective' (Hopfield, 1982) or 'cooperative' (Arbib, 1989, part 2) computation.

[25] Star studied the relationships between social metaphors and distributed AI. See: "The futility of the Turing Test comes not from lack of storage capacity or processing power, but from a fundamental misunderstanding of the nature of computers and society as closed, centralized and asocial . . . The simultaneous existence of multiple viewpoints and the need for solutions which are coherent across divergent viewpoints is a driving consideration of distributed artificial intelligence" (Star, 1989b, pp. 8-9).

2.3 The perceptron

In this section I look at Rosenblatt's 'Mark 1' perceptron, one of the most important early neural network machines. Afterwards, I discuss some of the most important results in early neural network research, namely Rosenblatt's and Widrow-Hoff's learning algorithms for single-layer neural networks (systems with only one layer of adjustable connections). These algorithms created considerable interest in neural network research, and encouraged other researchers to work in the field.

The Mark 1 machine (see photographs 1 and 2 in appendix 1) was the first important neural computer ever built. Its design was based on Rosenblatt's (1958a) perceptron system, and it was built and studied at Cornell Aeronautical Laboratories (CAL, Buffalo, New York). After completing his PhD in psychology at Cornell University in 1956, Rosenblatt went to CAL (today Arvin Calspan Advanced Technology Center) as a research psychologist. He later became chief of the 'Cognitive Systems Section.' In 1959 Rosenblatt went to Cornell University (Ithaca, New York), where he was first director of the 'Cognitive Sciences Research Program' and lecturer in psychology (for seven years), and then associate professor in the 'Division of Biological Sciences' (section of 'Neurobiology and Behavior') (see: New York Times, 1971).

Rosenblatt earned a reputation of making bold and risky choices, in research as well as in life. Rosenblatt died in a sailing accident in 1971.[26] This is how Richard O'Brien, then head of the Division of Biological Sciences at Cornell University and close to

---

[26] Rosenblatt died in an accident when sailing with two students in Chesapeake Bay, near Washington DC in his 43rd birthday, on the 11th of July 1971.

Rosenblatt, described Rosenblatt's approach to research after his death.

> "He would never do, as prudent professors do, pick out some small aspect of the physical universe which could be studied, and could reasonably be expected to produce solutions in a year or two so that someone could come up with a sensible reputation. Instead he would reach out and grasp the biggest problems that he could see, and apply himself, throw himself into the study of them, without any recognition of the fact that if he made bad choices, or if he chose, as he usually did, problems which were not likely to yield a solution within ten or twenty years, that this would redound to his disadvantage." (Congressional Record, 1971, p. 3)

The perceptron can be seen as one of Rosenblatt's bold choices, as it will be shown in this chapter and in the following one. Figure 2-4 below is a simplified diagram of Rosenblatt's perceptron machine. The perceptron can be defined as a single-layer feedforward neural network. In the case of the perceptron, 'single-layer' means that there is only one layer of modifiable connections (namely the connections from association units to response unit in figure 2-4), and not that it has only one layer of connections. As it is shown in figure 2-4, the perceptron had two layers of connections, namely the ones from sensory units to association units, and the ones from association units to the response unit. In later sections two more types of single-layer neural network machines will be studied: Widrow's adalines and madalines, and Minos 2, the machine built at Stanford Research Institute (SRI). All these machines were feedforward systems. In a feedforward neural computing machine, activation can only spread in the direction from input units to output units, and not in the opposite direction. In Rosenblatt's perceptron (see figure 2-4 below) activation spreads throughout the network in the following direction: sensory units → association units → response unit. One of the most important characteristics of both Rosenblatt's perceptron and Widrow's adaline were their training algorithms. These techniques guarantied that if a single-layer

network was physically capable of embodying a solution to a problem (i.e. of classifying a set of inputs into some predefined categories), then it could learn it after a finite number of training steps (what this means will be seen later).

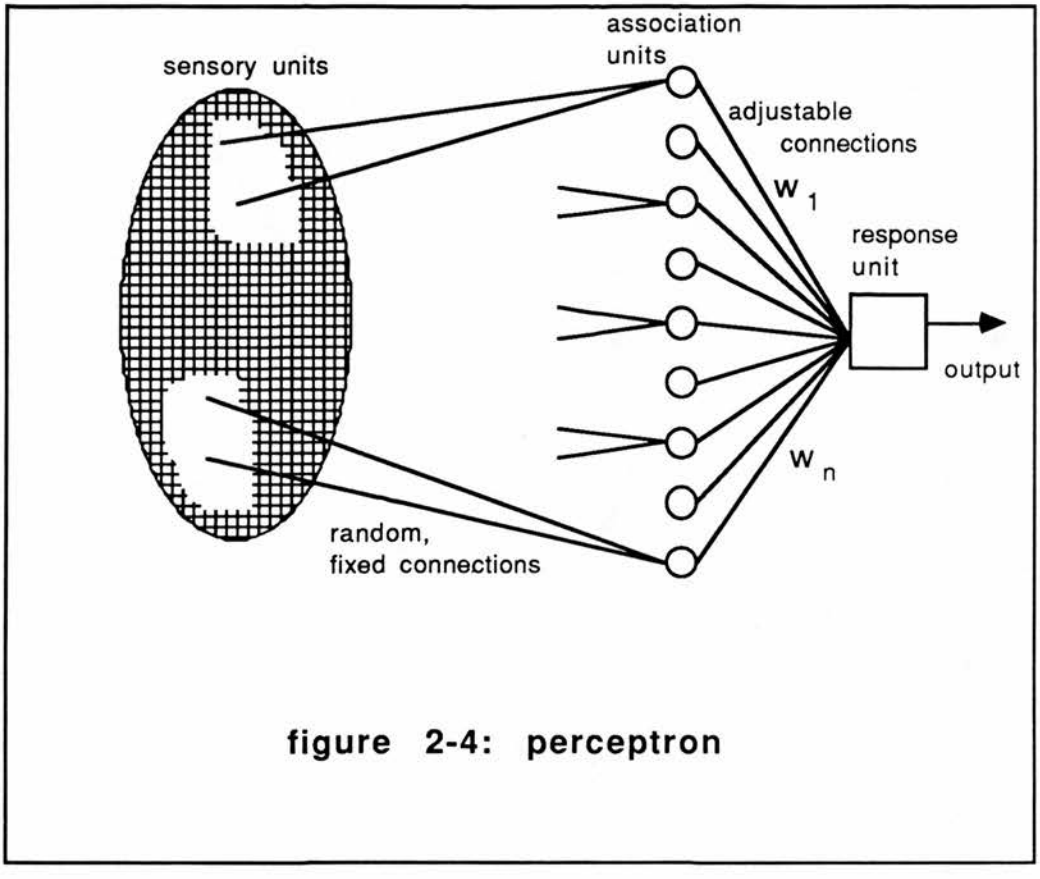


figure 2-4: perceptron

The perceptron of figure 2-4 above has three types of processing units or 'neurons,' namely sensory units, association units, and response units. Sensory and association units fire if their input signals equal or exceed their threshold values. Sensory units have got only one 'input line,' and consequently the computation they perform is just to detect whether there is an external input (in their little square of the retina in the perceptron) or not. The connections between sensory units and association units have a fixed value in the perceptron.

The perceptron built by Rosenblatt's group at CAL had eight response units, but only one of them is represented in figure 2-4 above. The response units had modifiable incoming connections.

Units of this type have been the building block of many neural computing systems since then. The response units of the perceptron were basically McCulloch and Pitts formal neurons with modifiable connections. These processing units (figure 2-5 below represents one of them) performed typically the following computation:

$$o_j = 1 \quad \text{if} \quad \sum_{i=1}^{n} v_i \, w_{ji} \geq \theta_j$$

$$o_j = 0 \quad \text{if} \quad \sum_{i=1}^{n} v_i \, w_{ji} < \theta_j$$

In words, a (response) unit (say unit j) fires (produces output activation, $o_j = 1$) if the sum of activation it receives from other units (v_i) equals or exceeds its threshold value (θ_j). Note that the activation which a unit receives through each of its input lines is multiplied by the value of those lines or connections (w_{ji} would be the value of the connection from unit i to unit j).

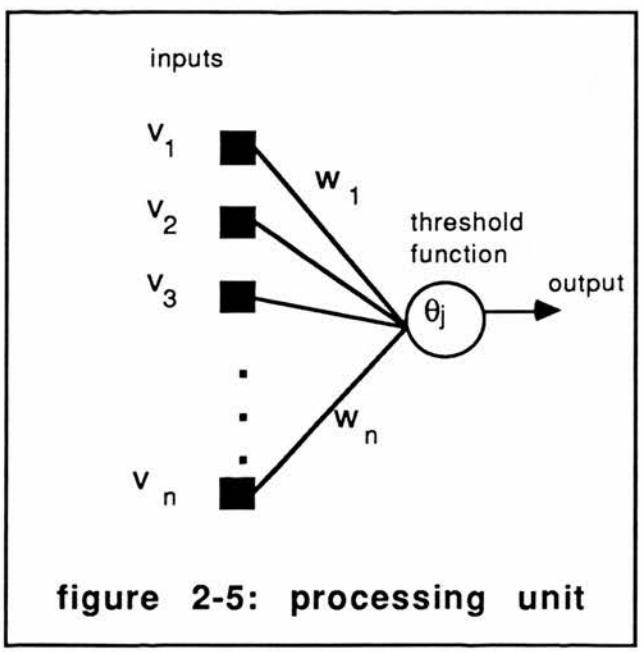


figure 2-5: processing unit

The type of neural network architecture represented in figure 2-4 is not the only one studied by Rosenblatt and his colleagues. They studied cross-coupled perceptrons (systems in which there are connections among units of the same layer) and

are connections among units of the same layer) and back-coupled perceptrons (systems in which connections in the direction 'sensory units ← association units ← response units' are allowed), as well as feedforward (or, in his terminology, series-coupled) perceptrons (see Rosenblatt, 1962a). Here I will concentrate on the single-layer feedforward perceptron (henceforth the perceptron), because it was the best studied and most successful of Rosenblatt and his group's neural network systems. Rosenblatt's work on more complex systems (e.g. on networks with more than one layer of adjustable connections or 'multilayer networks') was of a more 'programmatic ' character. Very importantly, it will be shown in later sections that Rosenblatt's (single-layer feedforward) perceptron was at the centre of a controversy which shaped the development of early neural network research.

One important aspect of the architecture of the perceptron was the pattern of connections from sensory to association units. I will be seen in later sections that some important arguments developed in the perceptron controversy were related to this issue. Figure 2-6 below (after Rosenblatt, 1962a, p. 139) shows one sensory-to-association connection pattern studied by Rosenblatt.

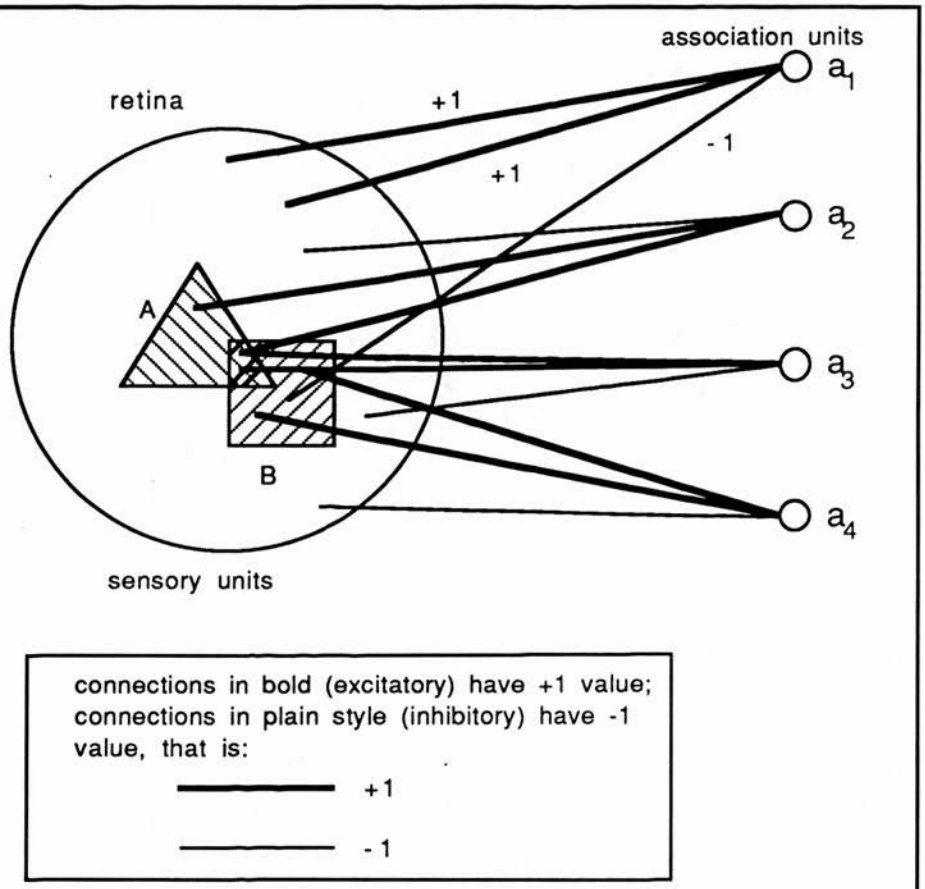


figure 2-6: connections from sensory to association units in the perceptron

As represented in figure 2-6, each association unit in this system receives connections from three sensory units. Another way of saying this is that this is a perceptron of 'order' three (Minsky & Papert, 1969). Usually the order of a perceptron was not too big, that is an association unit did not receive connections from many sensory units. On the other hand, the order was typically fixed, although Rosenblatt studied perceptrons with random orders too. In the Mark 1 perceptron built at CAL by Rosenblatt's team, an association unit could receive up to 40 connections from sensory units (Block, 1962, p. 142). In many of the systems studied by Rosenblatt (1962a), an

association unit typically received about ten connections from sensory units. These connections could be excitatory or inhibitory. In the perceptron of figure 2-6, each association unit (a_i) receives connections from three sensory units, two of them being excitatory (i.e. of +1 value) and one of them being inhibitory (i.e. of -1 value). In this example the thresholds of all the association units have value +2. It is also assumed that connection activate only when they 'detect' an input in the retina (e.g. the inhibitory connection to unit a_4 does not get activated, and therefore does not intervene in the summation done by a_4). Two stimuli appear in the retina of this perceptron, stimulus A (a triangle) and stimulus B(a square).[27] Association unit a_1 responds to none of them, since the activation it receives (-1) is smaller than its threshold (θ_1=2). Unit a_2 responds to the triangle, and not to the square. Unit a_3 responds to both stimuli, and unit a_4 responds only to the square.

The origin of the connections from sensory units to association units was typically random in Rosenblatt's machine.[28] Photograph 1 in appendix 1 shows the random connection patterns between sensory units and association units (see the left part of the machine). The issue of randomness was treated in section 2.2. On the one hand, this randomness was a consequence of Rosenblatt's inspiration from certain contemporary work in brain research (in particular the belief in the statistical character of the finer structure of neural connections). On the other hand, by using randomness Rosenblatt and his colleagues wanted to show that the perceptron, unlike the von Neumann computer, did not require a precise wiring pattern. I will compare this aspect of the perceptron with other early neural network machines later in this section and in the coming section.

---

[27] The fact that these stimuli are different geometrical figures is irrelevant here. Later in the dissertation I will come back to the point that the criterion used by the perceptron in classifying stimuli was not geometrical similarity, but amount of 'retinal' overlap.

[28] Later on Rosenblatt also analysed perceptrons with more constrained sensory-to-association connections, but the random character of the connections from sensory to association units was a distinctive property of the Mark 1 perceptron.

Now I will comment some other important details of the Mark 1 perceptron before coming to the issue of learning in single-layer neural networks. Mark 1 was built at Cornell Aeronautical Laboratories (CAL, Buffalo, New York) towards the end of the 1950s.[29] Rosenblatt's research group at CAL included R. D. Joseph and H. D. Block on the mathematical side, Carl Kesler, Trevor Barker, David Feign and Louise Hay on the digital computer programming side, and Charles Wightman (project engineer) and Francis Martin on the engineering side (Rosenblatt, 1962a, p. vii). Mark 1 had 400 sensory units (photosensitive receptors), 512 association units, and eight response units.

The analog weights (continuously variable quantities) of the association-to-response connections of the perceptron were implemented using motor-driven potentiometers.[30] In photograph 2 of appendix 1 Charles Wightman, Mark 1 perceptron project engineer, is holding a subrack of 8 motor/potentiometer pairs (each motor/potentiometer pair implemented one adaptive weight). These devices had an important limitation: their size. The 512-weight Mark 1 perceptron occupied a whole room at CAL (see photographs 1 and 2 in appendix 1). This meant that a machine with thousands (let alone millions) of modifiable weights would be impractically large. Rosenblatt was rather worried about the problem of analog weight implementation, and he tried to encourage other groups of engineers to produce more adequate solutions. In fact this is how the neural network project at Stanford Research Institute started, as it will be shown in section 2.4. At Stanford University, Widrow and his colleagues developed another technological solution, the so-called 'memistors.' These developments will be examined in section 2.4. It will seen that, important though the problem of

---

[29] There are CAL reports on the perceptron project from 1957 to 1960: (Rosenblatt, 1957), (Rosenblatt, 1958b), (Wightman, 1959), (Hay, 1960), (Hay, Martin, & Wightman, 1960), and (Rosenblatt, 1960). See: (Hawkins, 1961), (Block, 1962), and (Rosenblatt, 1962a).

[30] A technological solution similar to this was used by W. K. Taylor of University College, London, in a neural network-like pattern recognition machine he built in 1959 (Aleksander & Morton, 1990, p. 57).

implementing adjustable analogue weights was, it was not the only difficult problem that early neural network researchers had.

The question of learning was a very important one in Rosenblatt's perceptron and in other early neural network machines. A perceptron is not programmed in the sense of conventional computers. In order for a perceptron to improve its performance in some classification task, someone has to adjust its modifiable connections according to a rule (or learning algorithm). Two learning algorithms for single-layer neural networks were developed in 1960. One by Rosenblatt (1960) himself, and the other one by B. Widrow and M. Hoff (1960). These results created quite a lot of interest in neural network research at the time. Widrow (interview) said that the two algorithms were developed independently:

> "It turns out that the first publication that Rosenblatt had made of the perceptron learning algorithm was done, I think, in January or February 1960, and our first publication of the LMS [least mean square, or Widrow-Hoff] algorithm was done in June of 1960. Both of these works were done independently."

Rosenblatt's (1960; 1962a, ch. 5) 'perceptron convergence theorem' showed that, if a perceptron was physically capable of performing a classification task, then that perceptron could be 'taught' that task in a finite number of training cycles. A training cycle involves presentation of a pattern, observation of the output given by the machine, and adjustment of the connections according to an algorithm. In other words, a perceptron is capable of learning (in finite time) any classification which its architecture is capable of supporting. This result was important and created considerable excitement. Rosenblatt (1962a, p. 111) formulated the perceptron convergence theorem in the following way:

> "Given an elementary α–perceptron, a stimulus world W, and any classification $C(W)$ for which a solution exists; let all stimuli in W occur in any sequence, provided that each stimulus must reoccur in finite time; then beginning from

an arbitrary initial state, an error correction procedure . . . will always yield a solution to *C(W)* in finite time . . ."

An α–perceptron is a perceptron whose connections are adjusted according to the α algorithm (to be described below). The perceptron convergence theorem was proved for the simplified perceptron of figure 2-7 (representing the adjustable part of the original perceptron after removing the sensory-to-association fixed connections). The perceptron learning algorithm says that, for learning to occur, it is necessary that the perceptron architecture be capable of embodying the desired input/output classification. Proving whether a classification can be done by the simplified perceptron of figure 2-7 (let alone the Mark 1 perceptron with its first layer of randomly wired connections) is a NP-complete problem, that is it is exponentially intractable (the time it takes to solve it grows exponentially with the size of the problem). Thus although the perceptron rule is a powerful learning algorithm, training a single-layer neural network in a classification task is very much an empirical (experimentation-based) matter (where factors like the input/output training sample used and the generalisation abilities required after training are very important).

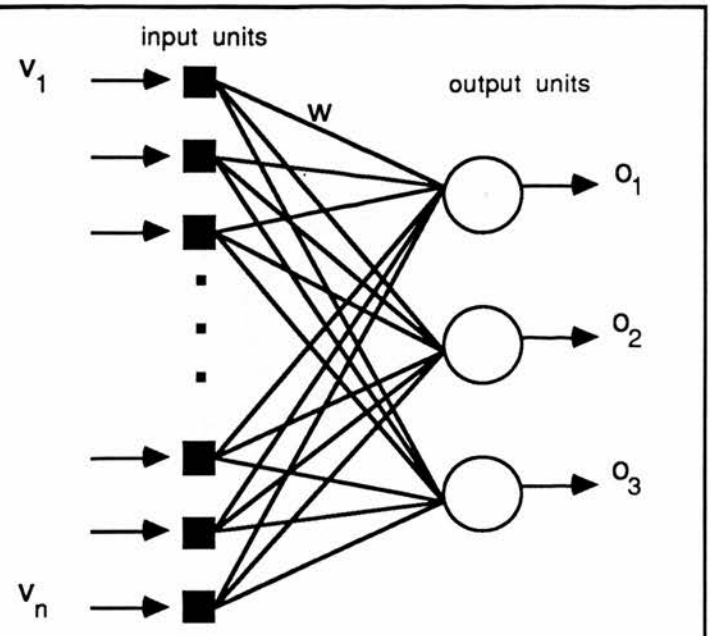


figure 2-7: simplified perceptron

For the simplified perceptron of figure 2-7, the α learning algorithm can be described as follows. After initialising the weights of the network (this involves setting the weights to small, random values), a binary input pattern (v_1, v_2, . . . v_n) is presented to the system. Activation spreads from input to output units, and the system produces an output or response (in this case the binary vector o_1,o_2, o_3). Each of the output units of the perceptron of figure 2-7 performs the same computation (summation plus thresholding) as the processing unit of figure 2-5. If the perceptron produces the desired output vector (i.e. $o_j=t_j$, being 't_j' the desired or target output vector) then no connection adjustment is carried out (i.e. $w_{ji}(\tau) = w_{ji}(\tau+1)$, being '$\tau$' an instant of time, and w_{ji} the connection weight from input unit i to output unit j). If, on the contrary, the actual output of the system is not the desired one, then connection modifications are carried out according to the following rule:

$$\text{if } o_j = 0, \text{ and } t_j = 1, \text{ then } \ldots w_{ji}(\tau+1) = w_{ji}(\tau) + \eta\ v_i$$

if $o_j = 1$, and $t_j = 0$, then . . . $w_{ji}(\tau+1) = w_{ji}(\tau) - \eta\ v_i$

η is the adjustment rate: $0 \leq \eta \leq 1$

In words, if the actual output of an output unit (unit j) is 0 when it 'should' be 1 then the value of the activated connections coming to unit j is increased (the quantity $[\eta v_i]$ is added to them). If, on the contrary, the output of an output unit (unit j) is 1 when it 'should' be 0, then the value of the activated connections ending in unit j is decreased (the quantity $[\eta v_i]$ is subtracted from them). So the weight modification can be defined as:

$$\Delta w_{ji} = \eta\ (t_j - o_j)\ v_i$$

In Rosenblatt's α–reinforcement rule only the weight value of active connections (w_{ji} if $v_i=1$) is incremented or decremented. Each weight is adjusted by the value of the input to that connection multiplied by the adjustment rate (η). If the input to a line is $v_i=0$, then $\eta v_i=0$, and therefore the value of the corresponding weight remains the same (that is $w_{ji}(\tau+1)=w_{ji}(\tau)$).[31] It is important to remember that weight modification in the perceptron depends on evaluation of performance by an external agent, and it is therefore called 'supervised learning.'[32] Another important characteristic of the perceptron learning algorithm is that error is minimised for each

---

31 The threshold of the output units can be represented by extra weights and bias inputs (see Beale & Jackson, 1990, ch. 3). Supposing a perceptron with only one output unit, the threshold of that unit can be represented by an extra weight (w_0) such that $w_0=-\theta$, and v_0 (the bias input) is always +1. Now the value of the threshold can also be modified according to the α algorithm. In the perceptron learning algorithm only the activated connections (i.e. connections for which $v_i=1$) are modified. The connection representing the threshold ($w_0=-\theta$) is always active, and therefore it will be modified. For example, if the output is correct, the activated connections are incremented, and therefore (because of the 'minus' sign) the threshold value is made smaller.

32 As opposed to other neural network learning schemes which do not involve a supervisor, such as competitive learning and what has come to be named 'Hebbian learning.' Hebb's (1949) notion of learning by synaptic strength modification was examined in section 2.1. Although the most influential learning algorithms of the 1960s belonged to the category of supervised learning, later researchers also used unsupervised learning. The term 'Hebbian learning' has come to be associated with one type of unsupervised learning (see Hecht-Nielsen, 1990, pp. 50-56, and Hertz et al., 1991, pp. 197-215).

output unit independently of the others. Widrow and Hoff's LMS algorithm, to be described below, was different in this respect. LMS minimised error as summed over all the output units.

At the same time as Rosenblatt was working on his perceptron learning algorithm, Bernard Widrow and his PhD student Marcian E. Hoff (1960) of Stanford University, Department of Electrical Engineering, developed another learning algorithm for single-layer neural networks. This algorithm is known as the Widrow-Hoff or LMS (least mean square) weight modification rule. Widrow and Hoff developed this algorithm for their 'adaline' artificial neuron. Figure 2-8 represents an adaline. In photograph 3 of appendix 1 Bernard Widrow is holding one of his adaline machines. The history of the term 'adaline' is curious. Originally, 'adaline' stood for 'adaptive linear neuron.'[33] But some years later, after the crisis of early neural networks (to be analysed chapter three), Widrow and his colleagues changed the name to 'adaptive linear element' (a tactical and rhetorical move!). Neural networks were no longer popular.[34]

---

33 "For the past several years, we at Stanford University have called this element 'ADALINE' (adaptive linear neuron)." (Widrow, 1962, p. 435)

34 See Widrow at the 1987 Snowbird, UT, conference, as quoted in (Anderson & Rosenfeld, 1988, p. 124).

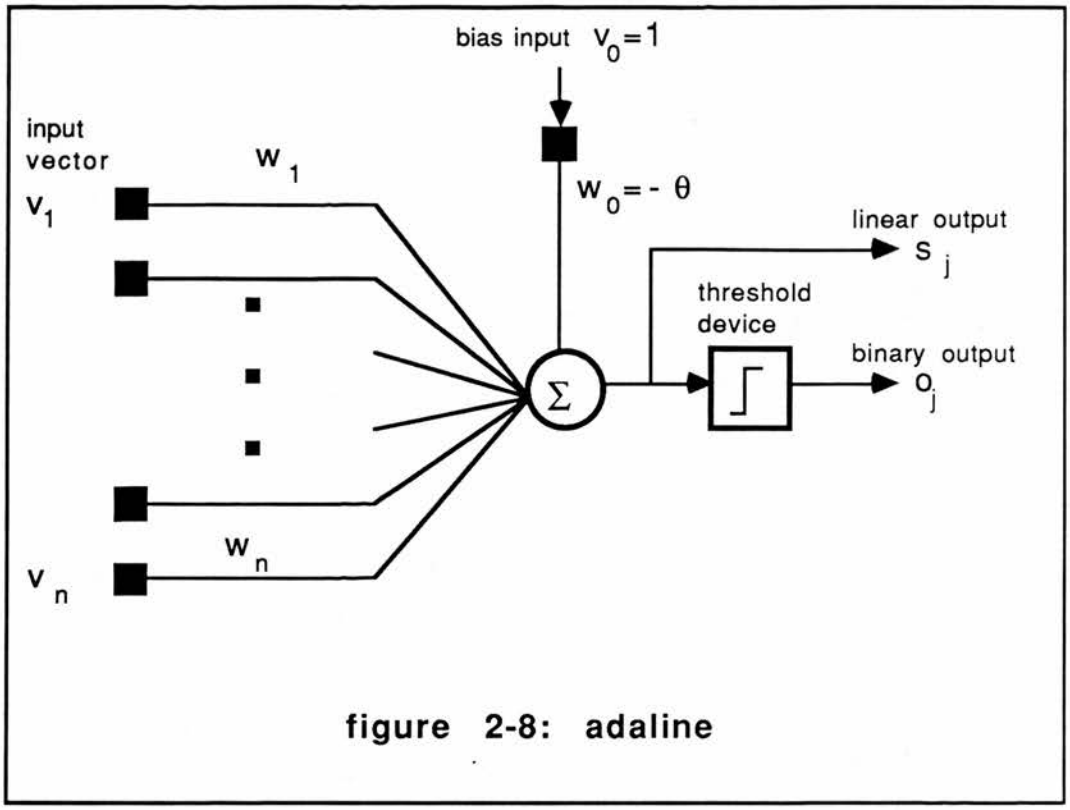


figure 2-8: adaline

As figure 2-8 shows, the adaline 'neuron' has several input units connected to one output unit. These connections are modifiable. In the adaline the threshold is treated as a weight ($w_0=-\theta$), which is always 'on,' that is it receives a bias input which is permanently activated ($v_0=1$). An adaline can be compared with each of the output units of the simplified perceptron of figure 2-7. There are, however, significant differences too. One important difference between the output units of the simplified perceptron of figure 2-7 and the adaline is that in the adaline there are two kinds of outputs, namely a linear output (the weighted sum of activation) and a binary output (the activation resulting from the thresholding function) (see figure 2-8). In the perceptron weights are adjusted according to the difference between the binary output and the desired output (for that unit), whereas in the adaline adjustment is made according to the difference between the *linear* output (s_j) and the desired output (see figure 2-8). In the adaline weight adjustment is proportional to the amount of error made, whereas

amount of error made, whereas in the perceptron algorithm the value of the adjustment is always the same.

Weights in the adaline are adjusted as follows. Supposing that the 'neuron' has already been initialised, and that a pattern 'p' has then been presented, the system then produces a linear output (s_j) and a binary output (o_j, which is the actual classification performed by the machine). If the classification is correct, no adjustment is carried out. But if the output produced by the machine is not the desired one, then the error made (d) is calculated as follows:

$$d_j = t_j - s_j$$

Thus the error (d_j) is the difference between the desired output (t_j) and the linear output (s_j). Assuming that the units have binary activation values, weight adjustment in the adaline can be described as follows.[35]

$$w_{ji}(\tau+1) = w_{ji}(\tau) + \eta\ d_j(\tau)\ v_i(\tau)$$

$0 \leq \eta \leq 1$ (being η the weight change rate, like in the perceptron algorithm)

'τ' is an instant of time. Note that the combination of d (error) and v (input activation) controls weight change. The error 'd' controls the sign of the weight change, positive (increment) or negative (decrement).

For a network composed of more than one adaline (similar to the perceptron of figure 2-7) Widrow and Hoff's algorithm minimises (mean-squared) error averaged over all the training patterns and summed over all the output units. Remember that the perceptron algorithm minimises error for each output unit independently. In a system of more than one adaline, like the one in figure 2-7, the LMS algorithm "minimizes the squares of the differences between the actual and the desired output values summed over

---

[35] See (Beale & Jackson, 1990, p. 50).

the output units and all pairs of input/output vectors" (Rumelhart, Hinton, & Williams, 1986, p. 322-324).[36]

For input/output pattern p, LMS minimises the following error measure:

$$E_p = \frac{1}{2} \sum_j (t_{pj} - o_{pj})^2$$

The total error or cost function which Widrow and Hoff's algorithm minimises will then be the sum of all the errors over all the training input/output patterns: $E = \Sigma E_p$. Widrow and Hoff's algorithm is especially important because, as it will be seen in section 5.2, the back-propagation learning algorithm, one of the most successful in the re-emergence of neural network research in the 1980s, was developed as a generalisation of Widrow and Hoff's LMS algorithm.

Widrow and his colleagues developed their own technology for implementing adjustable analogue weights. The weights and threshold of the adaline of figure 2-8 and photograph 3 (in appendix 1) were implemented by Widrow and his colleagues using their own technology, the so-called 'memistors,' for 'resistors with memory' (Widrow, 1960, 1962).

> "A memistor consists of a conductive substrate with insulated connecting leads, and a metallic anode, all in an electrolytic plating bath. The conductance of the element is reversibly controlled by electroplating . . ." (Widrow, 1962, p. 457)

---

[36] Widrow and Lehr (1990, pp. 1423–1424) recently described the properties of LMS algorithm in the following terms: "The minimal disturbance principle [adapt to reduce the output error for the current training pattern, with minimal disturbance to responses already learned] . . . was the motivating idea that led to the discovery of the LMS algorithm [Widrow & Hoff, 1960] . . . In fact, the LMS algorithm had existed for several months as an error reduction rule before it was discovered that the algorithm uses a instantaneous gradient to follow the path of steepest descent and minimize the mean-square error of the training set . . . The LMS algorithm corrects error, and if all input patterns are of equal length, it minimizes mean-square error. The algorithm is best known for this property."

In an adaline one memistor was used to implement each adaptive connection, plus one more to implement the threshold (ibid., pp. 456-457). Memistors were much smaller than the motor/potentiometers of the perceptron (memistors were about 1 cm long). The adaline (see photograph 3 in appendix 1) was the size of a small suitcase. Widrow and Hoff founded a company called 'Memistor Corporation' (Mountain View, California) with a view to developing and commercialising their machines and devices. They did not sell many adalines, but memistors were quite popular at the time for a variety of uses (Widrow, interview).[37] Motor/potentiometers and memistors were not the only devices developed for implementing adjustable weights in early neural network research. Researchers at Stanford Research Institute used multi-aperture magnetic cores in their Minos machine (to be described in section 2.4).

Rosenblatt's and Widrow-Hoff's learning algorithms created quite a lot of interest in neural network research. It was shown, for the first time, that a neural network could 'learn,' that is it could improve its performance without having to be programmed (in the von Neumann computer sense) for doing so. It was also shown that some kind of (very primitive) 'cognitive' or 'intelligent' behaviour could emerge from a non–symbolic substratum. Rosenblatt's perceptron convergence theorem received much attention at the time (for a review of proofs for the perceptron convergence theorem published at that time, see [Nilsson, 1965, ch.5]). Among the people who were impressed by Rosenblatt's result was Marvin Minsky, who later became one of the most influential critics of neural networks. In an interview carried out by Jeremy Bernstein (1981, p. 99) in 'The New Yorker,' Minsky expressed 'admiration' for Rosenblatt for having developed the perceptron convergence theorem:

---

37 "We were selling adalines, we didn't sell very many, not enough to support the business. But what was quite popular at the time was the memistor. People were using that for all sorts of uses in electronic circuits. It's the equivalent of a transistor with a built-in integrator, it's a mixture of an integrator and a transistor, all in one unit." (Widrow, interview)

". . . 'Rosenblatt made a very strong claim [Minsky's words], which at first I didn't believe,' Minsky told me. 'He said that if a Perceptron was physically capable of being wired up to recognize something, then there would be a procedure for changing its responses so that eventually it would learn to carry out the recognition. Rosenblatt's conjecture turned out to be mathematically correct, in fact. I have a tremendous admiration for Rosenblatt for guessing this theorem, since it is very hard to prove . . .' . . ."

Although Rosenblatt's and Widrow-Hoff's learning algorithms created quite a lot of interest in neural network research and encouraged a considerable number of researchers to work in the field, many problems remained. For example, the architecture of the single-layer perceptron could not realise certain classification tasks (which means, of course, that these tasks could not be learnt). And, even if the perceptron was 'physically capable' of realising a classification task, other important questions remained, such as how long it would take, which kind of training sample had to be used, etc.. The problems of the perceptron will be discussed in section 3.2.

In this section I have examined Rosenblatt's perceptron, and I have discussed two of the most important results of early neural network research, namely Rosenblatt's and Widrow-Hoff's algorithms for connection adjustment in single-layer neural networks. In the coming sections I will look at other aspects of single-layer neural network machines. In chapter three I will show that single-layer neural networks were at the centre of a controversy which had very important consequences in the evolution of neural network research.

2.4 The Madaline and Minos projects

In the previous section I examined Frank Rosenblatt's perceptron and certain parts of the work carried out by Bernard Widrow and his colleagues. In this section I look at some aspects of the two other main projects of early neural network research: the Madaline and Minos projects. The Madaline machine was built and studied by Widrow and his colleagues at Stanford University, Department of Electrical Engineering (Stanford, California). The Minos machine was built and studied by Charles Rosen and colleagues at Stanford Research Institute (Menlo Park, California). I will look at the Madaline project first, and then at the Minos project. I will also examine one important difference between the perceptron (see section 2.3) and the machines studied in this section, namely that neither Widrow and colleagues nor the SRI researchers used random connections in their machines. In section 3.2 I will examine other aspects of the perceptron, the Madaline, and Minos, focusing on the problems that early neural network researchers were having with their machines.

For Widrow and his colleagues at Stanford University the adaline (see section 2.3) was only the first step in their neural network project. Adalines ('adaptive linear neurons') were used as the building blocks of more complex machines. Widrow and his colleagues M. E. Hoff, T. M. Cover, R. J. Brown and W. C. Ridgway built and studied machines composed of many adalines, the so-called 'madalines' (multiple adalines). Figure 2-9 (after Widrow, 1962, p. 445; and Widrow & Lehr, 1990, p. 1426) represents one of the madalines studied by Widrow and his colleagues.

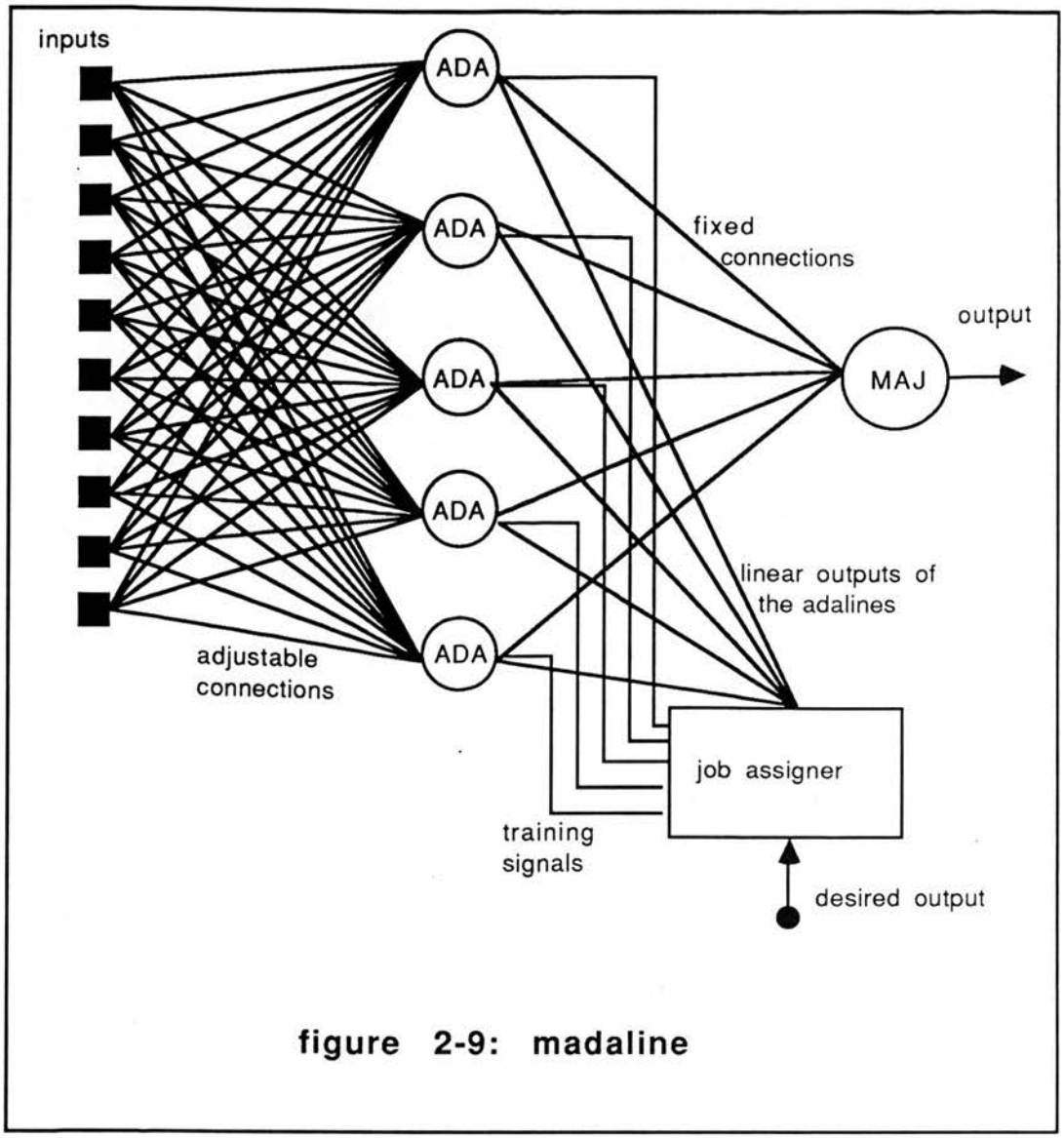


figure 2-9: madaline

In the madaline in the figure above there are two layers of connections. The connections from input units to adalines ('ADA's in the figure) are modifiable, but the connections from the adalines to the 'majority' logic device ('MAJ' in the figure) have got a fixed value. The activation values used in this machine by Widrow and colleagues were bipolar (-1, +1). The majority function or majority-vote-taker element produces an output of +1 if more than half of the (fixed) connections it receives are positive. Widrow and his colleagues used other functions too in this fixed part of the system, such as 'and' and 'or.'

The madaline can be compared with the adaline in terms of pattern storage capacity. Widrow and colleagues showed that the average number of random patterns which could be stored in (or learned by) an adaline is approximately equal to twice the number of its weights (Widrow, 1962). T. M. Cover (1964) and R. J. Brown (1964), two postgraduate researchers of Widrow, confirmed this in their dissertations (Widrow & Lehr, 1990, pp. 1418-1419). Widrow and his colleagues found that the average number of patterns which could be stored in a madaline was approximately equal to twice the number of adaptive weights (Widrow, 1962, p. 446). Thus both madalines and adalines had approximately the same capacity per adaptive weight, but the madaline had many more weights and could perform more complex classifications.

In the early 1960s Widrow and his PhD researchers M. E. Hoff (1962) and W. C. Ridgway (1962) studied several rules for adapting the connections of the madaline (Widrow & Lehr, 1990, p. 1420). One way of adapting weights in the madaline of figure 2-9 was the following (Widrow, 1962, pp. 444-446). Supposing that the desired response is +1, and three of the five adalines (ADAs) produce output -1 (and therefore MAJ produces -1 as the system's output), then one of those three adalines has to be adapted. The adaline adapted is the one whose output is nearest the desired output. The weight modification process is done by the job assigner (see figure 2-9 above). The idea was to adapt the minimum number of adalines necessary to produce the correct response. Adaptation was carried out according to Widrow and Hoff's LMS algorithm. In his 1962 dissertation, Ridgway found that, if the weights of the madaline were capable of embodying a given classification, then a solution could usually be found by modifying the connections of the madaline according to the LMS procedure (Widrow, 1962, pp. 444-446; Nilsson, 1965, pp. 99-101).

The biggest madaline machine built by Widrow and colleagues in the 1960s contained 1000 adaptive weights:

> "In the early 1960s, a 1000-weight Madaline was built out of hardware and used in pattern recognition research. The weights of this machine were memistors, electrically variable resistors developed by Widrow and Hoff which are adjusted by electroplating a resistive link (Widrow, 1960b)." (Widrow & Lehr, 1990, p. 1420)

In the early 1960s pattern recognition research was carried out using several versions of the madaline machine. Widrow and colleagues were aware of the limitations of this single-layer machine, and started to build and study madalines with two layers of adjustable connections (multilayer machines). In section 3.2 I will discuss some of these developments.

Also in the early 1960s, a group at Stanford Research Institute (SRI) started a project aiming at building and studying a machine — named Minos — which had a madaline-like architecture. This was the third main project of early neural network research (the other two were Rosenblatt's and Widrow's). The SRI researchers built the largest early neural network machine (in terms of number of adjustable weights).

The origins of the SRI neural network project go back to an 'unsolicited' personal visit by Frank Rosenblatt to SRI in the late 1950s. At that time Rosenblatt was travelling to several research centres throughout the United States, trying to convince people to carry out research on perceptron-like machines (Rosen, interview). Charles Rosen, one of the members of the SRI neural network group, described Rosenblatt's visit to SRI in the following terms:

> "Around 1959 or so, certainly not much later than that, we had an unsolicited visit from Frank Rosenblatt, who came around the country giving talks on what he called the perceptron. He had just begun to write some of the famous, well, I guess, pretty famous early papers on it. He was a psychologist with not much of a background in engineering. He knew some mathematics. And, really [laughing], in our first view of him, he was not very prepossessing: a short fellow, with very heavy glasses. . . But he had a deep voice,

> and when he started to talk about what he was doing your picture of him completely changed. He was an interesting man, a very interesting man. Later, as we got to know him better, he earned our deep respect" (Rosen, interview)

It was seen earlier that one of the problems of the perceptron was the excessive size of the analog weights. Rosenblatt and his colleagues at Cornell Aeronautical Laboratory had used motor-driven potentiometers in their Mark 1 machine. Rosenblatt thought that, in order to build larger machines, a different way implementing the weights was needed. Rosenblatt (1962a, p. 582) expressed the two main hardware implementation problems of the perceptron as follows:

> "The construction of physical perceptron models of significant size and complexity is currently limited by two technological problems: the design of a cheap, mass-produceable integrator [adjustable weight], and the development of an inexpensive means of wiring large networks of components. The Mark 1 . . . employs motor-driven potentiometers for integrators, and a large patch-panel for connections — both intolerable solutions for very large systems."

According to Charles Rosen (interview), Rosenblatt insisted on the problem of the implementation of adjustable weights when he visited SRI:

> "A. E. Brain and I both listened to Frank [Rosenblatt]. He said that he was looking for other people to do some work in perceptrons. In particular, he was looking for groups that knew how to make devices. What he wanted was somebody that would make a device that could serve as a neural element, a weight changing device. Well, he came to the right guys, because Brain and I had a background in physics and engineering, and a great knowledge of many devices, and we invented several such solid state devices and systems."

C. Rosen and A. E. Brain started to work informally on implementation aspects of perceptron-like machines. Then they first got funding for a small project on neural networks from the

Office of Naval Research (ONR). Later, in 1960, Rosen succeeded in getting funding for a neural network project of considerable size from the United States Army Electronics Command (formerly US Army Signal Corps, Fort Monmouth, New Jersey). The name of the project was 'Graphical data processing research study and experimental investigation' (I call it the Minos project here).[38] The Minos project was carried out from 1960 to 1965. Researchers who worked in this project include D. J. Hall, S. W. Miller, J. H. Munson. G. H. Burch, Richard O. Duda, H. Fossum and G. E. Forsen, as well as Rosen and Brain. On the theoretical side of the group, Nils Nilsson's (1965) work on neural networks was very important. Nils Nilsson, who later became a leading researcher in symbolic AI, joined the SRI group after Rosen, Brain, and their colleagues had already started their neural network project.

> "Some people at SRI, notably a man named Ted Brain [A. E. Brain] had an idea about how to implement these analogue, adjustable weights with multi-aperture magnetic cores. So to test that idea, the group at SRI was building a perceptron with these multi-aperture magnetic cores. That's about the time I joined the group. I was very interested not so much in multi-aperture magnetic cores or any other technology for implementing these things, but more on the basic ideas of whether they would be any good even after you implemented them." (Nilsson, interview)

Peter Hart, who later became well known for his work in pattern recognition with the above mentioned Richard Duda (Duda & Hart, 1973), joined the group later, when the project was about to be finished. Like in the other main early neural network projects, a good part of this group's research efforts was directed towards hardware implementation. They built the Minos machine, the biggest (in terms of number of weights) of the early period of neural computing (see photograph 4 in appendix 1).

---

[38] The first contract was from April 1960 to May 1963, DA 36–039 SC–78343, SRI Project 3192, and is summarised in report number 12, June 1963 (Brain & Munson, 1966, p. 1). The second contract, DA 36–039 AMC–03247(E), SRI Project 4565, from June 1963 to January 1966) was summarised in report number 22, by Brain and Munson (1966).

The implementation approach of much of early neural network research contrasts with the simulation (in digital computers) approach of early symbol-processing AI. There are two reasons for the importance of implementation in early neural network research. One is that digital computers were not widely available in the early 1960s; the other one is that the 'philosophy' of the neural network approach did not favour the use of the von Neumann computer (see section 2.2). Rosen (below) made some comments in the line of the first of those reasons.

> "At that time [in the beginning of the SRI's neural network project], we didn't have a computer, so we couldn't simulate the circuits of the perceptron. That came later. Other people could, but not even Frank [Rosenblatt] himself had the computer to do the simulation. So if you wanted to do experimental stuff, you had to build right from scratch all the bits and pieces." (Rosen, interview)

When Rosen says 'other people did have computers for simulation' he may be referring to the 'centres of excellence' of symbolic AI research, funded mainly by DARPA, namely MIT, Carnegie Institute of Technology (now Carnegie-Mellon University), and Stanford University. As James Fleck (1982, p. 208) pointed out, computer resources and funding for AI research in the 1960s were concentrated in these centres (I will come back to some aspects of this issue in section 2.5).

The actual architecture of the Minos machine built at SRI resembles much more Widrow's madaline than Rosenblatt's perceptron. Figure 2-10 (after Brain, Forsen, Hall, & Rosen, 1963, p. 10) is an schematic representation of the first version of the Minos machine.

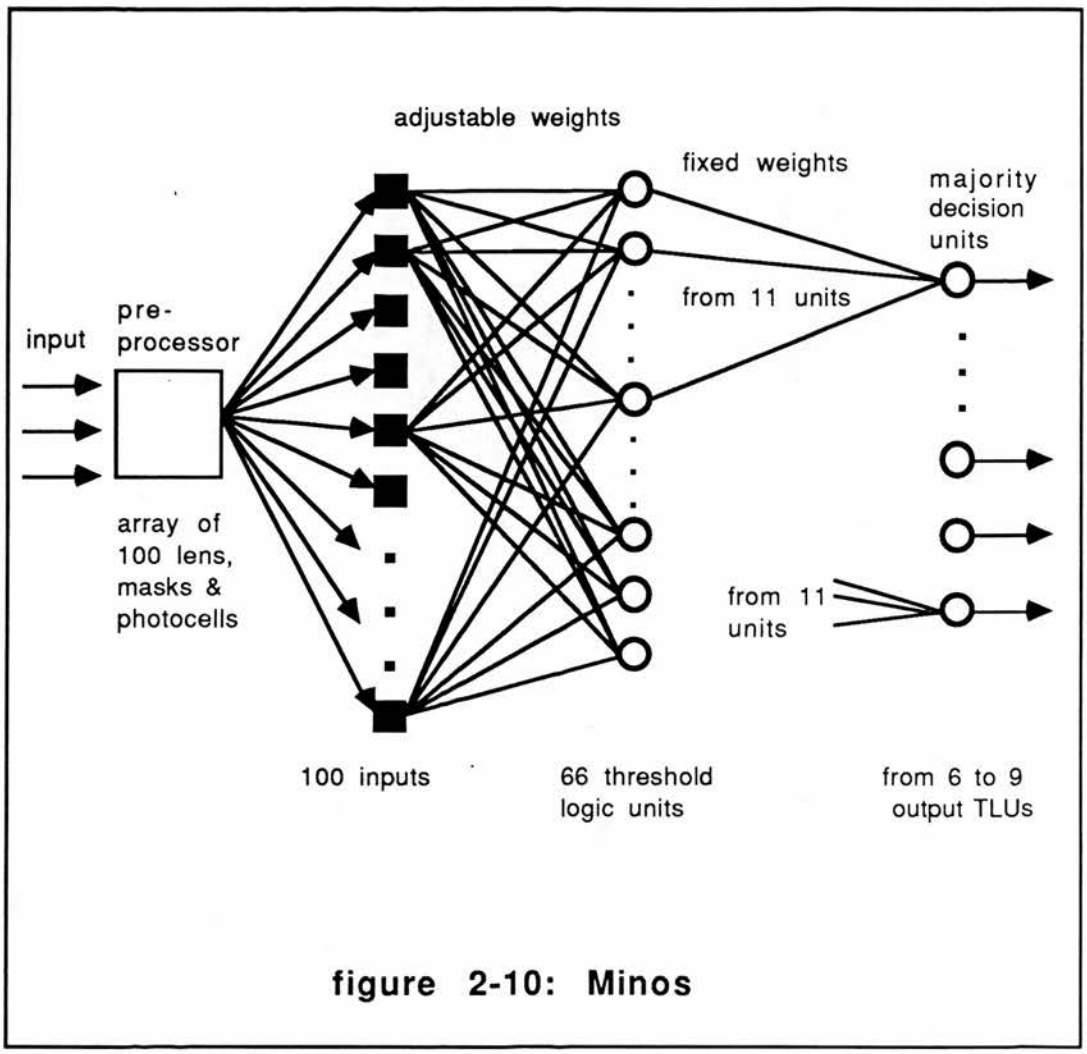


figure 2-10: Minos

The structure of Minos is divided into three main parts: a preprocessing system, a layer of adjustable weights, and a layer of fixed weights (these last two parts were similar to a madaline). Particular attention was devoted to implementing the adjustable weights and to the problem of preprocessing input images. The SRI researchers used multi-aperture magnetic cores (Brain, 1961) to implement the modifiable weights after rejecting other alternatives like electroplating (the one used earlier by Widrow).

> "Our early experiments with electrochemical and electroplating weights did not indicate much promise for real-time stability, consistent with low cost. We have, therefore, concentrated on magnetic, solid-state realizations. One of the first, devised by A. E. Brain (1961),

> utilizes multi-aperture, ferrite cores. Partial blocking of a minor aperture provides many stable levels of read-out voltage." (Brain, Forsen, Hall, & Rosen, 1963, p. 5)

The SRI researchers developed their own preprocessing system, which will be described below. Before that I will make a little digression to discuss some of the differences between machines like Minos or the Madaline and Rosenblatt's perceptron. There is nothing like the perceptron's random pattern of sensory-to-association connections in the Minos or madaline machines. Neither Widrow and colleagues nor the researchers at SRI used random connections in their machines. I asked both Bernard Widrow and Charles Rosen (of the SRI group) about this issue. Widrow and his colleagues built their machines from a more engineering approach (in contrast with Rosenblatt's stronger interest in the brain), and they saw no advantage in using random connections. Widrow said that he could not convince Rosenblatt of the disadvantages and problems created by the random organisation of the first layer of connections of the perceptron.

> "I took that little machine [pointing at the 'adaline;' see photograph 3 in appendix 1] to Rosenblatt's laboratory, to Cornell University, to visit Frank [Rosenblatt] one day. He had this Mark 1 perceptron, several racks of equipment, and I did a little problem in front of him and his students. He tried to do that with the perceptron, with the hardware that they had. They had a hell of a time. This tiny thing! [the adaline] The perceptron is actually much more than the adaline, it had all this nonlinearity in the first layer. I was telling him, 'what is the point in having all that nonlinearity?' He said that he believed that that was the way nature wired up the human retina, all randomly connected. Nobody believes that today, but he believed that at that time and that is why he felt that the perceptron had to have all that. So I shrugged my shoulders, because I knew that all that random stuff from the front serves no purpose. If there is any structure in the pattern, it breaks it all up, and makes the recognition task much more difficult. But I could never convince him." (Widrow, interview)

In their research on the preprocessing part of Minos, Rosen and his colleagues at SRI were inspired by the work done by Hubel, Wiesel, and others on early vision in the cat and in the frog. This work showed the importance of localisation in early visual processes in the visual cortex of the cat and in the optic nerve of the frog, and was carried out by D. H. Hubel and T. N. Wiesel (1959) and J. Y. Lettvin, H. R. Maturana, W. S. McCulloch and W. H. Pitts (1959) respectively. Later, in 1981, Hubel and Wiesel (from Harvard University) shared the Nobel Prize with Roger Sperry for their contributions to neurophysiology. It is important to note that the research at the basis of the design of the Mark 1 perceptron (Rosenblatt, 1958a) was carried out before Hubel and Wiesel and Lettvin and his colleagues published their results.

Inspired by this work by Hubel-Wiesel and by Lettvin and colleagues, the SRI researchers did not use random connections, and dedicated a significant research effort to the preprocessing part of their Minos machine.

> "Rosenblatt's original perceptron idea was that the first layer accepted information in an encoded form — or from a camera, or what have you — randomly. There were random connections into the second layer [of units], completely random, and no preprocessing was done whatsoever. Of course, that's not the way we humans do it. At that time, the biologists, some who won the Nobel Prize [referring to Hubel, Wiesel, Lettvin, Maturana, etc.; the first two won the Nobel Prize in 1981], were studying the frog, the cat, and other small animals. The people that were studying the cat's eyes [Hubel and Wiesel] found that the first few layers of neurons of the cat's visual cortex were doing a lot of preprocessing to extract lines and areas, and things that moved. They preprocessed an enormous amount of information, and selected certain features, which then went on to another layer, and then another layer, and so on, for processing. We watched this work very carefully, and said, 'we should do that too.' So we improved on the original perceptron organisation by doing preprocessing at the front edge. This is where we advanced work well beyond Frank Rosenblatt's." (Rosen, interview)

Figure 2-10 above shows the first version of Minos' preprocessor. It works as follows (Brain et al., 1963; Huber, 1967). An optical input image is presented to the machine, and 100 replicas of the input image are produced using 100 lenses. These 100 replicas are passed through a plate containing 100 photographic masks. Each mask acts as a feature filter. There is a photocell for each image-mask combination. Each photocell generates an analog signal proportional to the integral of the light transmitted. Each of these signals is then thresholded to produce 100 binary inputs to the adaptive part of the machine.

Later a second version of the preprocessor of Minos was built. It included a television camera instead of the original slide projector at the front of the preprocessing system. There was an array of 32x32 lens (1024 replicas of the input image) instead of the original 100. This offered the possibility of replicating each of the original 100 masks approximately 10 times in different orientations and translations using a plate with 1024 masks. 1024 analog signals were produced with the photocell system mentioned before, and were then thresholded to produce 1024 binary signals. Subsequently, these 1024 binary signals were reduced to 100 binary signals as the input to the adaptive (madaline-like) part of the machine. Each group of masks representing the same object was combined in an 'or' circuit to produce these 100 binary inputs. In the adaptive part of the machine, there were 100 input units fully connected to 66 threshold logic units (the usual neural network units). These 6600 adjustable weights were implemented using multi-aperture magnetic cores, as it was mentioned earlier.

Minos had from six to nine output units (or output threshold logic units, TLUs in figure 2-10) depending on the different versions. In the first version there were six output units. The 66 intermediate units were connected to these six units in groups of eleven, i.e. each output unit received connections from 11 intermediate units. The connections between intermediate and output units were fixed, and the computation performed by each

of the output units was a 'majority decision,' like in Widrow's madaline.

Brain and Munson (1966, pp. 27-56) and Huber (1967) described a number of experiments carried out with Minos for the classification of the parity function, military map symbols, and aerial photographs of tanks.[39] Since the arrival of a SDS 910 digital computer to SRI in June of 1964 (very late in the evolution of the Minos project), the performance of Minos 2 was compared with results obtained in digital computer simulations (Brain & Munson, 1966, p. 30).

Reports on the experiments carried out with Minos are quite incomplete; by reading them it is very difficult to assess Minos 2's performance. Swaine (1989b) confirmed this. William A. Huber (1967), Minos project monitor for the Army, and Brain and Munson (1966, pp. 27-30), reported on an early operational test carried out with the 100-element optical preprocessor version of Minos. The task was to classify 75 map symbols (each one mounted on a single 35 mm. slide). These symbols belonged to 15 categories, and each symbol was replicated 5 times with small translations or rotated. The output was coded using 9 bits. After 75 iterations (by 'iteration' the SRI researchers meant presentation of all input patterns and modification of weights), the number of errors per iteration was still considerable (about 5), but 'training was terminated because of a mechanical fault in the projector' (Huber 1967, p. 4). In this test Minos' performance was only measured with patterns from the training set (generalisation is always worse).

In the conclusions to their experiments, Brain and Munson (1966, p. 56) emphasise one particular problem, namely that of 'finding techniques to deal with overlapping figures' — the connectedness problem, which was later studied by Minsky and Papert (1969) (see section 3.3). Overall, the results obtained by carrying out

---

[39] The parity function consists of saying whether the number of input units activated after the presentation of a stimulus is odd or even. The parity function was analysed in detail in Minsky and Papert's (1969) study of Rosenblatt's perceptron.

experiments with Minos (as the ones obtained with the Madaline) do not seem to have been too encouraging. In section 3.2 I will discuss some of the limitations and problems of these machines.

In this section I have examined some aspects of the Madaline and Minos neural network machines. It has also been showed that neither Widrow and his colleagues nor the SRI researchers used random connection patterns in their machines. This was a considerable difference between Rosenblatt's perceptron and the other early single-layer neural network machines. In section 3.2 I will discuss the problems that early neural network researchers were having with their machines. It will be seen that these problems became increasingly worrying, and that by the mid-1960s early neural network research was in a situation of crisis.

2.5 The emergence of symbolic artificial intelligence

In this section I look briefly at some aspects of the emergence of symbolic AI. A detailed analysis of the emergence and development of symbolic AI falls out of the scope of this dissertation.[40] Here I examine some of the main ideas guiding symbol-processing AI, and I discuss some of the characteristics of its emergence and institutionalisation in the late 1950s and early 1960s. I conclude by pointing out that the claim that there were no 'credible alternative approaches' was used by symbol-processing AI researchers as a rhetorical tactic for legitimising their approach. In chapter three I will study the importance of the emergence of symbolic AI in the development of the perceptron controversy.

In the late 1950s and early 1960s, at the same time as early neural network researchers were working in their projects, the symbol-processing approach to the problem of studying cognition and building intelligent machines was emerging with increasing momentum. The emergence of symbolic AI was an important factor in the crisis of early neural network research (discussed in section 3.2) and in the development of the perceptron controversy (analysed in chapter two). Single-layer neural network machines had important problems of their own which researchers in the 1960s were far from solving satisfactorily. These included training multilayer networks and recognising objects as different from their background (the 'connectedness' problem, later studied by Minsky and Papert [1969]). Nevertheless, the emergence of an *alternative* approach to AI

---

[40] Historical accounts — developed from various viewpoints — of the emergence of symbolic AI have been written by authors including: Fleck (1978), McCorduck (1979), Fleck (1982), Newell (1983), Gardner (1985), and Pratt (1987). Fleck's (1978) study is especially interesting because of the amount of historical detail it offers. I am not aware of similar detailed studies covering later developments in AI research.

research was an important factor in the crisis of early neural computing in the 1960s and in the evolution of the perceptron controversy.

The process of emergence of symbolic AI in the United States can be placed in the ten years from 1955 to 1965. It is important to note that the end of this process and the crisis of neural network research happened at around the same time. This is not a mere coincidence: the emergence of symbolic AI was an important factor in the crisis of early neural networks and in the closure of the perceptron controversy. Symbol-processing AI defined intelligence in terms of symbol manipulation. Researchers within this tradition aimed at developing computer programs and computer systems which could carry out certain tasks (or solve certain problems) for which intelligence is necessary. Early developments in symbolic AI include research carried out by A. Newell, J. C. Shaw and H. A. Simon (chess-playing programs, the 'Logic Theorist' theorem proving program, the 'General Problem Solver'), the development of list-processing programming languages (like Newell, Shaw, and Simon's IPL languages and J. McCarthy's LISP), and Minsky's work on heuristic methods for problem solving. (Feigenbaum and Feldman [1963] is probably the most comprehensive compilation of early symbol-processing AI research.)

Some of the foundational elements of the symbolic approach, such as formal logic or the notions of computability and universality (in Turing's sense), were developed by the 1930s. The other fundational element, the von Neumann digital serial computer with its stored program and single central processing unit, was developed in the 1940s and 1950s, and became widely available towards the late 1960s. The early readiness of the foundational elements of symbolic AI was seen by Frank Rosenblatt (1962a, p. 21) as a disadvantage for neural computing:

> "While the monotypic [symbol-processing] approach arose rather suddenly with the advent of modern computers and control system theory, and rapidly advanced to a high level

> of mathematical sophistication, the genotypic [neural network] approach has been much more gradual in its development, and has not yet developed all of the mathematical tools required to deal adequately with its problems."

The digital computer became the experimentation tool of symbolic AI researchers. It will be seen later in this section that early symbol-processing AI researchers had privileged access to digital computer resources, unlike early neural network researchers. Within symbolic AI, emphasis was laid on the symbol-processing capabilities of the digital computer, rather than on the numerical ones. These researchers developed their own high-level programming languages, the most important of them being LISP (list processing) developed by McCarthy in 1960. In list processing languages, chain storage of data is used, which is rather different from the serial system of storage used in most numerical computer applications. Using serial data storage means that, in order to retrieve a datum, one needs to add its serial number to the start address. As V. Pratt (1987, pp. 224) pointed out, this is not the most adequate storage system for the manipulation of sequences of data (symbolic expressions) typical of symbolic AI:

> "For most numerical applications, this approach [serial data storage] works well, since what is usually required in such contexts is for a number to be taken from one location and have some operation performed upon it, and for the result to be stored. But non-numerical manipulations often require something different. Here the manipulation is typically not on a single datum but on a sequence of data: elements of the sequence are to be deleted, or moved, or extra elements added."

In list processing, data are stored in chains ('lists'). Each datum in a list has the address of the next datum in the list recorded on it, and this makes the insertion of new data — the manipulation of symbolic expressions — much easier.

> ". . . To insert a new datum near the beginning of a serial sequence means that all the data stored after the insert

> have to be moved one location on to make room. Programming is made easier . . . [in AI] if data is stored in chains. The first datum on a sequence to be stored is placed in a location: and next one or two locations are used to store an address, namely the address of the location in which the second datum of the sequence is to be stored. This location is in turn followed by the address of the third datum, and so on. Amending sequences is then much easier, involving the alteration of just a few addresses." (Pratt, 1987, p. 225)[41]

The clearest elaborations of the main theoretical assumptions of symbolic AI can be found in contributions including (Newell & Simon, 1976), (Newell, 1980), and (Newell, 1981). From a philosophical point of view Fodor's (1975) study is important in this respect. In cognitive science the assumptions of the symbol-processing approach were studied by researchers like Z. Pylyshyn (1980, 1984). Allen Newell and Herbert Simon's formulation of both the 'physical symbol system' and the 'heuristic search' hypotheses was especially clear. Newell and Simon (1976, pp. 41-42 and 51) argued that the two main assumptions of symbolic AI were (i) structure-sensitive symbol manipulation processes and (ii) heuristic search methods (to avoid combinatorial explosion of possibilities in problem solving):

> "*The Physical Symbol System Hypothesis.* A physical symbol system has the necessary and sufficient means for general intelligent action. By 'necessary' we mean that any system that exhibits general intelligence will prove upon analysis to be a physical symbol system. By 'sufficient' we mean that any physical symbol system of sufficient size can be organized further to exhibit general intelligence. By 'general intelligent action' we wish to indicate the same scope of intelligence as we see in human action: that in any real situation behavior appropriate to the ends of the

---

[41] One crucial point in languages like LISP is that: "computer programs themselves are represented in the machine as a sequence of codes, and that a list processing language like LISP makes it easy to treat programs themselves as manipulable data. Since the programs that are thus manipulable include the program that is doing the manipulation, we have here the facility to program recursively: to instruct a program to do things to itself . . . The power to program recursively means that all computable functions fall within the computer's scope" (Pratt, 1987, p. 226).

> system and adaptive to the demands of the environment can occur, within some limits of speed and complexity . . . *Heuristic search hypothesis.* The solutions to problems are represented as symbol structures. A physical symbol system exercises its intelligence in problem-solving by search — that is, by generating and progressively modifying symbol structures until it produces a solution structure."

Symbolic AI was composed of a variety of subareas including game playing, theorem proving, machine vision, natural language, cognitive modelling, robotics, machine vision, and heuristic programming. What united these subspecialties was their use computational tools, and their emphasis on the manipulation of symbolic expressions in a manner sensitive to their logico-syntactic structure. Cognitive processes were defined at the level of symbols (knowledge, concepts, plans, reasoning, problem solving, etc.), and not at the level of neurophysiology or at the neural network level. The digital computer (i.e. computer simulation) was the experimentation tool: computer programs (using high level languages like LISP) were constructed which could realise certain cognitive or intelligent tasks. In robotics research, symbol manipulation and knowledge representation were combined with vision and motor-control systems. More psychology-oriented AI researchers adopted the following methodology: first a piece of intelligent behaviour was described as accurately as possible; then computer programs which simulated that behaviour were designed and experimented with.[42] Rosenblatt opposed this type of methodology explicitly (see section 2.2).

Although in early symbolic AI some attention was devoted to learning (e.g. A. L. Samuel's checkers program) and pattern recognition, the 'relative weight' of these areas soon decreased. Instead, emphasis was given to questions of knowledge and

---

[42] Some symbolic researchers (like Newell and Simon) were more keen on the cognitive modelling aspects of AI than others (e.g. Minsky or McCarthy). "The psychological observations and experiments lead to the formulation of hypotheses about the symbolic processes the subjects are using, and these are an important source of the ideas that go into the construction of programs" (Newell & Simon, 1976, p. 49).

representation. This emphasis did not favour neural network research. In the important 'Semantic Information Processing' collection, Marvin Minsky (1968, pp. 17-18) stated that emphasis as a programmatic idea quite clearly, criticising at the same time neural networks:

> "Why do the heuristic programs solve much harder problems than do self-organizing systems? [this includes neural networks] Obviously, because they begin with methods appropriate to the classes of problems they are faced with, and because they are given enough specific factual knowledge about particular problems. They do not have to start from an unstructured basis to evolve everything they will need. But only to the extent that this 'knowledge' is suitably represented can the program's use of it be intelligent . . . We have agreed to set aside the problem of acquiring knowledge [i.e. learning] till we better understand how to represent and use it."

The process of emergence of symbolic AI in the United States was completed by the early or mid-1960s. J. Fleck studied the process of emergence and institutionalisation of symbolic AI and concluded that:

> "By the early 1960s, various successful programs had been written, resulting in a general air of optimism, and indeed by this time the paradigmatic structure of AI had been elaborated in essentially its complete form . . ." (Fleck, 1982, p. 178)[43]

Fleck (1982) showed that two of the main characteristics of the emergence and institutionalisation of symbolic AI were the influential role played by a rather small elite of researchers (and their students) belonging to a few research centres, and the concentration of computer resources (of basic importance in AI research) in a few centres of excellence. Allen Newell, Herbert Simon, John McCarthy, and Marvin Minsky were the core of that elite, and the main AI centres were the Massachusetts Institute

---

[43] The emergence of symbolic AI happened somewhat later in Britain, as Fleck (1982) pointed out. Pratt (1987) indicated that: "The year 1960 . . . we can take it as marking the end of its [symbolic AI's] emergence" (p. 235).

of Technology (MIT), Carnegie-Mellon University (formerly Carnegie Institute of Technology), and Stanford University.

The symbolic AI elite was rather successful both at defining the cognitive structure of their specialty and at gaining financial support from the US Department of Defense funding agencies, mainly from DARPA (the Defense Advanced Research Projects Agency). Fleck (1982) emphasised the importance of DARPA (at the time ARPA) in funding AI research projects since the 1960s, and the success of the symbolic AI elite in getting that funding.

> "The emergence of this group [McCarthy, Minsky, Simon, and Newell] as the establishment in AI was undoubtedly consolidated by their success in getting the backing of the United States Department of Defense, mainly through the Advanced Research Projects Agency (ARPA), which provided some 75% of United States AI funding for the ten years from 1964, and through the Air Force . . . Furthermore, the preference on the part of ARPA for concentrating resources in a few selected centres guarantied the position of the establishment, especially in view of the great expense of adequate computing facilities, which effectively barred other groups from competing" (Fleck, 1982, p. 181)

Fleck reports on ARPA funding for symbolic AI from 1964 onwards, but ARPA's support for that research started earlier. In Barber Associates' (1975, p. vi-53) study of ARPA from its creation in 1958 to 1974 it is indicated that the agency created a few centres of excellence in AI in the early 1960s. In ARPA's 'golden period' of funding, from 1961 to 1966, under directors J. P. Ruina, R. L. Sproull and C. M. Herzfeld, the agency supported heavily symbolic AI research at the mentioned centres of excellence. The general atmosphere at the agency in this period was one of emphasis on basic scientific research (ibid., p. i-9). Toward the end of the 1960s, coinciding with the Vietnam War, growth in ARPA funding stopped, and the pay-off of scientific research for military requirements was much more closely scrutinised. In section 3.4 I will come back to this issue when I discuss some aspects of funding for Rosenblatt's perceptron from the US military agencies.

It is important to note that, because of the concentration of computational resources in a few DARPA-funded centres of excellence in AI, elite symbolic AI researchers and their students had privileged access to digital computer facilities. Neural network researchers were not so lucky in this respect. For example, the SRI group did not get a computer for simulation experiments until 1964, very late in their neural network project.

As the process of emergence and institutionalisation of symbolic AI gained momentum, the crisis of neural network research (to be studied in section 3.2) deepened. It will be seen later (chapter three) that the emergence of symbol-processing AI played an important role in the development of the perceptron controversy. With an alternative approach backing their position, Minsky and Papert's criticism of perceptrons (and of neural networks in general) became more decisive. Furthermore, it will be seen that that criticism was aimed not only at showing the weaknesses of neural networks, but also the comparative strength of the symbol-processing approach.

'Eliminating' any alternative approaches was important for symbolic AI in its early years. So, to a certain extent, the crisis of neural networks helped legitimise the emergence and institutionalisation of symbolic AI (more on this issue in later sections). Some claims by E. Feigenbaum and J. Feldman (1963), the editors of the most important survey of early AI research, can be interpreted within this 'legitimisation context.'[44] Within this context, the claim that there was no credible alternative — and particularly the lack of credibility of neural networks as an alternative — was an important point in favour of the symbolic approach. Edward Feigembaum and Julian Feldman (1963, pp. v-vi) suggested this in their introductory remarks:

---

44 Minsky (1968, p. 8), Fleck (1978, p. 43), and Pratt (1987, pp. 235-236) all emphasise the importance of (Feigenbaum & Feldman, 1963) as the main collection of early AI work.

> "We have selected reports of research efforts which we feel outdistance all others in advancement towards this goal [artificial intelligence]. Such a criterion, as we see it, gives high priority to a particular brand of research, loosely labeled 'cognitive models.' An opposing school of thought, sometimes called 'neural cybernetics' or self-organizing systems,' has intrinsic fascination and has produced a considerable number of particular projects. Neural cybernetics approaches the problem of designing intelligent machines by postulating a large number of very simple information processing elements, arranged in a random or organized network, and certain processes for facilitating or inhibiting their activity. Cognitive model builders [i.e. symbol-processing AI researchers] take a much more macroscopic approach, using highly complex information processing mechanisms as the basis of their designs. They believe that intelligent performance by a machine is an end difficult enough to achieve without 'starting from scratch,' and so they build into their systems as much complexity of information processing as they are able to understand and communicate to a computer (using their programming techniques). The cognitive models approach has led to tangible progress (displacement toward the ultimate goal) in the field of artificial intelligence, while the progress to date in the neural cybernetics approach is barely discernible. On this basis, we feel that there is reason for our bias in favor of cognitive models . . ."

The first sentence from this quotation by Feigenbaum and Feldman means that papers on neural networks were excluded from 'Computers and Thought,' the most important collection of papers of early AI research. Arbib (1987, p. 7) confirmed this. Two pattern recognition papers (which were not too far from the neural network tradition) were accepted, but in the years after that the relative weight of pattern recognition within symbolic AI decreased significantly.[45]

---

[45] Fleck (1978, p. 47) pointed out that the 1973 biannual conference on AI decided not to accept papers dealing with pattern recognition. But the demise of pattern recognition within symbolic AI had started much earlier.

The opposition between symbolic AI and neural networks, which had both originated in the 1950s within the common 'umbrella' of cybernetics, developed into open controversy the late 1950s and early 1960s. I study that controversy in chapter three.

In this section I have examined some early developments in the symbol-processing approach to AI, and I have discussed some aspects of the emergence and institutionalisation of that approach in the late 1950s and early 1960s. In chapter three, I will study the importance of the emergence of symbol-processing AI in the crisis of early neural networks and in the perceptron controversy. I will show that certain researchers favouring the symbolic approach played a very important role in the controversy about single-layer neural networks (and neural networks in general).

◆ THREE

The Perceptron Controversy

3.1 The heat of the controversy

In this section I study some aspects of the controversy about Rosenblatt's perceptron (and neural networks in general) of the late 1950s and 1960s. First I make some comments about the sense in which I use the term 'rhetorical tactics.' Afterwards I study some of the rhetorical tactics used by both sides of the perceptron controversy. Rosenblatt's claims about the perceptron outside the research community, as well as inside, were the catalyser of the controversy. Rosenblatt's views about the capabilities of the perceptron were contested heavily by researchers who favoured the symbol-processing approach to AI, and a heated controversy developed. After examining the debating tactics used by researchers in favour and against the perceptron, I discuss the beginning (early 1960s) of Minsky and Papert's (1969) study. Minsky and Papert's aim was to show the limitations of the perceptron in a detailed and decisive way. In other words, Minsky and Papert aimed at making a decisive move in the controversy. The allocation of funding resources by the funding agencies was among Minsky and Papert's motivations to embark on their project. Minsky and Papert's move is analysed in detail in later sections.

I am using the term 'rhetorical tactics' or 'debating tactics' in a wholly non-pejorative sense, following authors such as Bruno Latour (1987) and Susan Star (1989a). The view in this thesis (see chapter one) is that scientific knowledge is generated and validated through controversies, and that controversies are solved (closed) through processes of rhetoric and gathering of allies and 'actants' (Latour, 1987).

Harry Collins (1981a) introduced the idea of 'rhetorical tactics' in his controversy/closure scheme for the social study of science. Collins used terms like 'non-scientific tactics' (in

inverted commas) to refer to debating or rhetorical tactics. But even though it is useful to place the words 'non-scientific tactics' in inverted commas, by using these words there is a risk that the phrase 'rhetorical tactics' may be interpreted as implying the intrusion of 'non-scientific' factors into science. The following is an example of Collins' (1985, pp. 143 and 152) use of the term 'non-scientific tactics:'

> "The knowledge which emerges from a core set is the outcome of an argument that may have taken many forms not normally viewed as belonging to science . . . Some 'non-scientific' tactics *must* be employed because the resources of the experimenter alone are insufficient . . . Once the scientific truth is known it is forgotten that non-experimental and 'non-scientific' negotiating tactics were necessary if closure was to be attained."

The phrase 'non-scientific tactics' in quotations like this might be interpreted as implying that these tactics do not really belong to science. But, of course, the whole of the recent sociology of science (including Collins' own work) cuts against the use of any dichotomy between 'scientific' and 'non-scientific' tactics. In order to avoid this risk of confusion, I have decided to use terms like 'rhetorical tactics' and 'debating tactics' here.

In this section I will show that rhetorical tactics of a great variety were used by the two sides involved in the perceptron controversy of the late 1950s and early 1960s. Here I will examine some of the most 'spectacular' rhetoric used. By 'spectacular' I mean colourful or lively, and of course I do not mean that later, more 'serious' moves in the controversy (such as Minsky & Papert's [1969] project, to be analysed in section 3.3) were not rhetorical or tactical. Quite the opposite: they were (using Latour's notion) stronger rhetorical moves. The move was not from rhetoric to 'truth,' but from weaker rhetoric to stronger rhetoric.

In short accounts of the history of neural network research, such as those appearing in short introductions to papers or books on

neural computing, it is often assumed that Rosenblatt made 'exaggerated claims' about the capabilities of his perceptron machine, and that the strong reaction of symbolic AI researchers against the perceptron (and neural network research in general) was 'justified' because of that. It is often assumed that Rosenblatt's claims had, at the end, a damaging effect upon the development of neural network research in the 1960s. The following quotation from Cruz (1988, p. 2) is an example:

> "In 1957, Frank Rosenblatt proposed a very influential neural net model called the 'Perceptron'. Great expectations were laid on this self-organizing system; in fact, these claims and expectations were overblown (a danger to be avoided at all costs by current researchers). This overselling caused a lengthy setback for the neural nets field. In 1968, Minsky and Papert published a book called Perceptrons, pointing out some of the limitations of that model. What ensued was an almost total shift in research funding from neural nets to the nascent field of Artificial Intelligence which was being defined by Minsky, McCarthy, Newell, Simon and others. The resulting 'dry spell' for neural nets lasted until the early 1980s."

In this chapter I will show that the (often made) assumption that Rosenblatt made exaggerated claims about the perceptron and that, because of that, the reaction by symbolic AI researchers against neural networks was 'justified,' misses important points, and therefore it is of little use in a study of the history of neural network research. In section 3.2 it will be seen that Rosenblatt and other neural network researchers acknowledged (and sometimes even insisted on) the limitations of their systems. Of course, rhetoric was used — and heavily — by both sides of the controversy. Indeed, it had to be used if the controversy was going to be resolved.

The perceptron controversy increased as Rosenblatt's perceptron project received considerable attention outside the research community. Rosenblatt (1958a, 1958b) published his first important papers on the perceptron in 1958. At that time, a team of researchers started to build the Mark 1 perceptron at Cornell

Aeronautical Laboratory (CAL) (see section 2-3) funded by the Office of Naval Research (ONR). In July 1958, the perceptron project was announced at a press release held in Washington DC. Marshall Yovits from ONR and Frank Rosenblatt participated in it. At that press release ONR announced its financial support for Rosenblatt's perceptron project. Rosenblatt made a demonstration of the capabilities of the perceptron simulating it in a digital computer. The perceptron project was widely reported in the press (e.g.: New York Times, 1958a, New York Times, 1958b, Newsweek, 1958, The New Yorker, 1958). The New York Times newspaper reported on the press release in the following terms:

> "The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence . . . Later Perceptrons will be able to recognize people and call out their names and instantly translate speech in one language to speech and writing in another language, it was predicted." (The New York Times, 1958a, p. 25: 2)

> "The concept of the Perceptron was demonstrated on the Weather Bureau's $ 2,000,000 IBM 704. In one experiment, the 704 computer was shown 100 squares situated at random either on the left or on the right side of a field. In 100 trials, it was able to 'say' correctly ninety–seven times whether the square was situated on the right or left. Dr. Rosenblatt said that after having seen only thirty to forty squares the device had learned to recognize the difference between right and left almost the way a child learns . . . Later Perceptrons, Dr. Rosenblatt said, will be able to recognize people and call out their names. Printed pages, longhand letters and even speech commands are within its reach. Only one more step of development, a difficult step he said, is needed for the device to hear speech in one language and instantly translate it to speech or writing in another language." (New York Times, 1958b, p. iv 9: 6)[46]

---

[46] Rhetoric used by Rosenblatt includes the following: "The question may well be raised at this point of where the perceptron's capabilities actually stop" (Rosenblatt, 1958a, p. 110). ". . . For the first time, we have a machine which is capable of having original ideas" (Rosenblatt, 1959, p. 449).

The experiment reported by The New York Times in the above quotation in which the perceptron learns between right and left is a good example of the capabilities of the perceptron (more on this in section 3.2). The perceptron was distinguishing between right and left, but it was not recognising the squares as the same object. If the objects presented had been triangles and squares situated on the right or on the left at random (being the triangles and squares of approximately the same size), the perceptron would not have been able to recognise the difference between squares and triangles. The similarity criterion used by the perceptron was the amount of overlap in its retina, and not the geometric similarity between the objects presented to it (more on this in section 3.2). But the fact that the perceptron learned something *at all* was presented as a success at the Washington press release.

In an article in the 'The New Yorker' the perceptron was compared with the 704 IBM digital computer in which the simulations of the 1958 press release were carried out.

> "Having told you about the giant digital computer known as IBM 704 and how it has been taught to play a fairly creditable game of chess, we'd like to tell you about an even more remarkable machine, the perceptron, which, as its name implies, is capable of what amounts to original thought. The first perceptron has yet to be built, but it has been successfully simulated on a 704, and it's only a question of time (and money) before it comes into existence. This about-to-be marvel is a lot more subtle than the 704; indeed, it strikes us as the first serious rival to the human brain ever devised, and our brain is thoroughly dazzled by the things it's said to do." (The New Yorker, 1958, pp. 44-45)

When people talk (and symbolic AI researchers did, as I will show below) about Rosenblatt making exaggerated claims they may be referring to this type of reporting in the press. The aim of the 1958 ONR press release was to announce and publicise the perceptron project in front of the wider public. This type of move

belongs to a context of legitimisation of scientific research in the wider society, and it is always a necessary move in science. The topic of the relationship between AI research and the wider society is a fascinating (and colourful) one, because there are many discourses and ideologies about human intelligence and human action in society which may feel 'affected' by AI's methods and conclusions. For a study of the images of AI in the wider society and the relationships between AI and ideologies, see Fleck (1984). But what interests me here is that the interaction with the wider society is always an element of scientific research. Furthermore, in Latour's terms, the bigger the 'inside' of scientific research, the bigger the 'outside;' in other words, the bigger a research project (the more allies and resources it needs), the more 'work' has to be done outside the laboratory. Bruno Latour (1987, p. 156) expressed this nicely:

> "Technoscience has an inside because it has an outside . . . The bigger, the harder, the purer science is inside, *the further outside other scientists have to go* . . . If you get inside the laboratory . . . you see science isolated from society. But this isolation exists only in so far as other scientists are constantly busy recruiting investors, interesting and convincing people. The pure scientists are like helpless nestlings while the adults are building the nest and feeding them. It is because . . . the boss . . . [is] so active outside that the . . . collaborator . . . [is] so much entrenched inside pure science."

The 1958 ONR press release belongs to the 'outside' of Rosenblatt's perceptron project, and the press clippings quoted above are a consequence of those 'outside' developments. But that press release was also a part of the 'outside' of ONR's involvement in funding scientific research projects. They also, as a government organisation, had to publicise and justify their activity. Marshall Yovits was responsible for the funding of the perceptron project at ONR (Information Systems Branch), and he participated in the Washington 1958 press release. When I interviewed him, he complained firmly about the reaction from symbolic AI researchers to Rosenblatt's work (this reaction is

studied below) and, importantly, to ONR's involvement in supporting that work.

> "Many of the people at MIT felt that Rosenblatt primarily wanted to get press coverage, but that wasn't true at all. As a consequence many of them disparaged everything he did, and much of what the Office of Naval Research did in supporting him. They felt that we were not sufficiently scientific, and that we didn't use the right criteria. That was just not true. Rosenblatt did get a lot of publicity, and we welcomed it for many reasons. At that time, he was with Cornell Aeronautical Laboratory, and they also welcomed it. But at ONR — as with any government organisation — in order to continue to get public support, they have to have press releases, so that people know what you are doing. It is their right. If you do something good, you should publicise it, leading then to more support." (Yovits, interview)

Rosenblatt's perceptron project's 'inside' was quite big, and so the 'outside' had to be quite big too. This outside was ONR, and ONR as a government organisation had to publicise and legitimise its projects in the wider society too. One interesting aspect of researchers' moves towards the outside is that, although these moves are necessary, researchers often complain about their consequences. Rosenblatt (1962a, p. v) complained about some of the press coverage of his perceptron machine:

> ". . . Reasons for the negative reactions to the [perceptron] program . . . [One of them] was the handling of the first public announcement by the popular press, which fell to the task with all the exhuberance and sense of discretion of a pack of happy bloodhounds. Such headlines as 'Frankenstein Monster Designed by Navy Robot That Thinks' (Tulsa, Oklahoma Times) were hardly designed to inspire scientific confidence."

It is interesting to make a brief digression here to compare the coverage of the 1958 perceptron press release with a report in the magazine 'Life' (Darrach, 1970) about the Shakey robot project carried out at Stanford Research Institute (SRI). This was the (symbolic AI) robotics project that the formerly neural

network group at SRI started after the end of their Minos project (see section 3.2). Darrach's article contained comments like these:

> "Marvin Minsky . . . recently told me with quiet certitude: 'In from three to eight years we will have a machine with the general intelligence of an average human being. I mean a machine that will be able to read Shakespeare, grease a car, play office politics, tell a joke, have a fight. At that point the machine will begin to educate itself with fantastic speed. In a few months it will be at genius level and a few months after its powers will be incalculable.' . . . In the interests of efficiency, cost-cutting and speed of reaction, the Department of Defense may well be forced more and more to surrender human direction of military policies to machines that plan strategy and tactics. In time, say the scientists, diplomats will abdicate judgement to computers that predict, say, Russian policy by analyzing their own simulations of the entire Soviet state and of the personalities — or the computers — in power there." (Darrach, 1970, pp. 58d and 66)

The SRI researchers criticised this article heavily, and Minsky denied the quotations attributed to him (McCorduck, 1979, 234-235). This move is similar to the above mentioned one by Rosenblatt. First, scientists have to move to the outside, but afterwards they criticise some of the (sometimes inevitable) consequences of those moves. Darrach's article contained controversial claims and exaggerations, but it was widely read (McCorduck, 1979, p. 235) and, it seems to me that it was beneficial for the SRI researchers in terms of getting funding. Charles Rosen, one of the researchers from the SRI group, gave me his view of the episode:

> "A writer came [to SRI], and interviewed me, and also interviewed Minsky, and other people in AI. Then he wrote an article [Darrach, 1970]. There was some good stuff in it, but he also wrote much rubbish . . . Minsky got very mad . . . Shakey was described pretty well, but there was also a lot of rubbish. Anyhow, it was a good article in some ways. We got a lot of notoriety from it!" (Rosen, interview)

After this short, 'comparative' digression, let me come back to the reactions to Rosenblatt's perceptron project and the 1958 press release. Researchers in favour of the symbol-processing approach to AI reacted strongly against Rosenblatt's claims. Pamela McCorduck (1979, p. 87) documents some of the rhetoric used in the controversy in her historical study of symbolic AI:

> "As time went on, the perceptron began to acquire a certain amount of notoriety. Besides its simplicity, there was another reason for its growing fame, and that was Frank Rosenblatt himself. Present day researchers remember that Rosenblatt was given to steady and extravagant statements about the performance of his machine. 'He was a press agent's dream,' one scientist says [McCorcuck does not disclose the name], 'a real medicine man. To hear him tell it, the Perceptron was capable of fantastic things. And maybe it was. But you couldn't prove it by the work Frank did' . . ."

Several of the 'debating tactics' studied by Susan Star (1989a, pp. 88ff and 134ff) in her historical study of localisationist research can be found in the perceptron controversy too. In the quotation above symbolic AI researchers used tactics like sarcasm and caricature: 'a press agent's dream,' 'a real medicine man,' 'to hear him tell, his machine was capable of fantastic things.'

A frequent rhetorical move by Rosenblatt's critics was to accuse him of 'irritating people.' This can be seen as an appeal to the authority and respectability of symbol-processing AI, as compared with the lack of authority (and respect for the *status quo*) of the view of their opponent (i.e. Rosenblatt).

> "Case-Western's Leon Harmon, who worked on the von Neumann machine at the Institute for Advanced Studies at Princeton, and who describes himself as perhaps the first computer operator, still seethes about walking into the Smithsonian and discovering that beside the von Neumann machine, which well deserved to be there, stood a Perceptron, sharing floor space as if it were equally important. Harmon doubts that we'll ever learn much about

> brain operation from studying electronic hardware, and believes that the really interesting and potent things the computer in our head does are inscrutable . . . Rosenblatt only irritated him." (McCorduck, 1979, p. 88)

But, of course, opinions were divided (otherwise there would not have been a controversy!).

> " . . . 'He *did* irritate a lot of people,' says W. W. Bledsoe of the University of Texas speaking of Rosenblatt, 'but he also charmed at least as many, and I count myself among them. Just when you were thinking that Frank [Rosenblatt] didn't have another trick up his sleeve, along he'd come, and he'd be so darn convincing, you know, he just had to be right' . . ." (ibid., p. 88)

One can infer Rosenblatt's charisma from the quotation above. The controversy of the perceptron was often personalised in the figures of Rosenblatt and Minsky. They were the 'symbolic leaders' or spokesmen of their respective positions. The rivalry between them may perhaps go back to their high school time. Frank Rosenblatt and Marvin Minsky knew each other well since they were teenagers. They had been classmates at Bronx High School of Science, New York (Bernstein, 1982, pp. 58-61; McCorduck, 1979, p. 87).[47] Minsky and Rosenblatt engaged in heated debates at scientific conferences in the late 1950s and early 1960s. Both McCorduck and Berstein reported on this:

---

[47] Minsky's description of the atmosphere at Bronx High School of science as it appears in Bernstein's (1981, pp. 58-61) 'New Yorker' article is interesting in this respect: ". . . In 1941 . . . Minsky entered the Bronx High School of Science. Bronx Science had been created just three years before to attract and train young people interested in the sciences. (Two of the 1979 Nobel laureates in physics — Steven Weinberg and Sheldon Glashow — were classmates at Bronx Science in the late forties, along with Gerald Feinberg, who is now the chairman of the physics department at Columbia, and during their senior year there the three taught themselves quantum mechanics.) 'The other kids were people you could discuss your most elaborate ideas with and nobody would be condescending,' Minsky said in recalling the experience there. 'Talking to people in the outside world was always a pain, because they would say, 'Don't be so serious — relax.' I used to hate people saying 'Relax.' I was a hyperactive child — always zipping from one place to the other and doing things very fast. This seemed to bother most adults. But no one at Science felt that way. Later, when I went to Harvard, I was astonished at how much easier the course work was there than it had been at Science . . . Frank Rosenblatt, who was tragically drowned in a boating accident in 1971, was also one of my classmates at Science' . . ."

> "Another who was irritated by Rosenblatt was Marvin Minsky, perhaps because Rosenblatt's Perceptron was not unlike the neural-net approach Minsky was alternately intrigued and frustrated by. Many in computing remember as great spectator sport the quarrels Minsky and Rosenblatt had on the platforms of scientific conferences during the late 1950s and early 1960s." (McCorduck, 1979, p. 88)

> ". . . Rosenblatt's claim was that after a finite number of adjustment the machine would learn to recognize patterns. Rosenblatt was an enormously persuasive man, and many people, following his example, began to work on Perceptrons. Minsky was not among them . . . Minsky and Rosenblatt engaged in some heated debates in the early sixties. During my discussions with Minsky, he described what the issues were. 'Rosenblatt made a very strong claim [Minsky speaking], which at first I didn't believe [referring to the perceptron convergence theorem; see section 2.3] . . . Rosenblatt's conjecture turned out to be mathematically correct, in fact . . . However, I started to worry about what such a machine could *not* do." (Bernstein, 1981, pp. 96-99)

I will come back to Minsky's 'worries about what the perceptron could not do' in section 3.3. Charles Rosen, who worked in the Stanford Research Institute (SRI) neural network project in the early 1960s, made the following comment about the debates between Minsky and Rosenblatt:

> "I think that there was a surmise that Minsky and others had not gotten anywhere in their early work with neural nets and here was somebody [Frank Rosenblatt], an upstart, working on neural nets, and getting some fame, and getting a lot of press — Frank got a lot of press at the time — and Minsky was very upset about a field he had abandoned due to little success." (Rosen, interview)

I infer from my personal contact with some of the neural network researchers who participated in the perceptron controversy (see list of those interviewed in appendix 2) that the atmosphere was rather bitter at times, and that there where moments when diplomacy was left behind, and researchers engaged in 'open confrontation.' Minsky and Rosenblatt were the

spokesmen, but controversy extended to other neural network projects, such as Rosen and colleagues' Minos project at SRI.

> "Minsky and his crew . . . thought that Frank Rosenblatt's work was a waste of time, and they certainly thought that our work [at SRI] was a waste of time . . . Minsky really didn't believe in perceptrons, he didn't think it was the way to go . . . I know he knocked the hell out of our perceptron business." (Rosen, interview)

The existence of heated debate was confirmed by Rosenblatt (1962a, p. v) himself in his main book, 'Principles of Neurodynamics:'

> "That the aims and methods of perceptron research are in need of clarification is apparent from the extent of the controversy within the scientific community since 1957, concerning the value of the perceptron concept."

Later in the same section Rosenblatt refers to Minsky in diplomatic terms as the 'loyal opposition' (but not without some irony: see the term 'entertaining' below):

> "See, for example, Minsky (1961) ['Steps toward artificial intelligence'], for an entertaining statement of the views of the loyal opposition, which includes an excellent bibliography." (ibid., p. vi)

Approximately at the time when Seymour Papert went to MIT, in 1963, Minsky and Papert decided to develop further their views about the limits of the perceptron with a view to publishing a book about the subject (this move by Minsky and Papert is studied in section 3.3). Seymour Papert was a South African mathematician who had been working on the question of learning with Jean Piaget in Geneva for several years (Bernstein, 1981, pp. 99-100).

Minsky and Papert's decision to start a detailed study of what perceptrons could and (mainly) could *not* do was motivated in part by the competition for funding resources between symbolic AI and neural networks. According to Minsky, neural network

researchers (presumably referring to Rosenblatt among others) 'were trying to get money to build bigger machines.' Minsky and Papert felt that they had to do something to stop that:

> "In the late 1950s, after Rosenblatt's work, there was a great wave of neural network research activity. There were maybe thousands of projects in the early 1960s, after Rosenblatt's book [presumably referring to Rosenblatt (1958a,1958b)]. For example Stanford Research Institute had a good project. But nothing happened. The machines were very limited. So I would say by 1965 people were getting worried. They were trying to get money to build bigger machines, but they didn't seem to be going anywhere. That's when Papert and I tried to work out the theory of what was possible for the machines without loops [feedforward perceptrons]." (Minsky, interview)

Seymour Papert (1988, pp. 4-5) too admitted recently the importance of the funding issue in the controversy:

> "There was *some* hostility in the energy behind the research reported in *Perceptrons* . . . Part of our drive came, as we quite plainly acknowledged in our book, from the fact that funding and research energy were being dissipated on . . . misleading attempts to use connectionist methods in practical applications."

The 'heat' of the controversy is reflected in Minsky and Papert's (1969, pp. 18-20) account of the reasons why they decided to elaborate their views about the perceptron further in the form of a comprehensive study of the limitations of perceptron.

> "The popularity of the perceptron as a model for an intelligent, general purpose learning machine has roots, we think, in an image of the brain itself as a rather loosely organized, randomly interconnected network of relatively simple devices. This impression in turn derives in part from our first impressions of the bewildering structures of the brain . . . The mystique surrounding such machines is based in part on the idea that when such a machine learns the information stored is not localized in any particular spot but is, instead, 'distributed throughout' the structure of the machine's network. It was a great disappointment, in

> the first half of the twentieth century, that experiments did not support nineteenth concepts of the localization of memories (or most other 'faculties') in highly local areas . . . In this setting, Rosenblatt's (1958a) schemes quickly took root, and soon there were perhaps as many as a hundred groups, large and small, experimenting with the model . . . The results of these hundreds of projects and experiments were generally disappointing, and the explanations inconclusive . . . The machines usually work quite well on very simple problems but deteriorate very rapidly as the tasks assigned to them get harder . . . Both of the present authors (first independently and later together) became involved with a somewhat therapeutic compulsion: to dispel what we feared to be the first shadows of a 'holistic' or 'Gestalt' misconception that would threaten to haunt the fields of engineering and artificial intelligence as it had earlier haunted biology and psychology."

The 'more scientific than you' tactic is used here by Minsky and Papert, who use the adjective 'mystique' to refer to neural network machines with distributed memory. Minsky and Papert appeal to authority against the 'holistic threat.' The heat of the controversy is also reflected in expressions like 'therapeutic compulsion,' and 'misconception that would threaten to haunt AI.' See also the rhetoric employed by Minsky when J. Bernstein (1981, p. 100) interviewed him:

> "In the middle nineteen-sixties Papert and Minsky set out to *kill the perceptron*, or, at least, to establish its limitations — a task that Minsky felt was a sort of *social service* they could perform for the artificial-intelligence community." (emphasis added)

There is an interesting similarity between Minsky and Papert's move to intervene in the perceptron controversy decisively and some developments of the controversy about the detection of gravitational radiation of the 1970s as studied by Collins (1981b, pp. 47-48):

> "Another scientist said . . . that Quest [imaginary name] embarked on this as a sort of a *holy crusade* . . . In sum, it

> can be said with some degree of certainty that Quest and his group *set out to kill* Weber's findings in the shortest possible time . . . They did their experiment with the intention of developing a position from which they could more effectively destroy Weber's findings . . . Quest acted as though he did not think that the simple presentation of results with a low key comment would be sufficient to destroy the credibility of Weber's results. In other words, he acted as one might expect a scientist to act who realized that simple evidence and arguments are not sufficient to settle unambiguously the existential status of a phenomenon" (emphasis added)[48]

Collins (1981b, 1985) pointed out that, in situations of controversy, the scientists he studied behaved 'as though' they thought that the simple presentation of scientific results was not enough for the closure of the debates. But in using phrases like 'as though' there is a risk of separating scientific and 'non-scientific' factors, as I said in the beginning of this section. Using that phrase, one could say that Minsky and Papert acted 'as though' they thought that the simple presentation of scientific evidence showing the limitations of perceptrons was not enough to close the controversy. The risk is interpreting this 'as though' expression as implying that that evidence was *really* enough to close the controversy. And if this implication is accepted, then the *result* of the closure of the controversy is treated as the *cause* of that closure (which is, of course, contradictory). Is scientific change is understood as a social process of controversy and closure, then rhetoric is an essential element of scientific activity, and therefore it is better not to use terms like 'as though' or 'non-scientific tactics.' The move in science, as Latour (1987) suggested, is not from rhetoric to 'truth,' but from weaker to stronger rhetoric. I will come to the 'stronger rhetoric' in section 3.3.

In order to show the extent of the perceptron controversy, it is interesting to repeat some of the rhetorical expressions that were used in it: 'many remember as great spectator sport the

---

48 See also (Collins 1985, p. 95).

quarrels Minsky and Rosenblatt had;' 'Rosenblatt irritated a lot of people;' 'Rosenblatt was given to steady and extravagant statements about the performance of his machine;' 'Rosenblatt was a press agent's dream, a real medicine man;' 'to hear Rosenblatt tell it, his machine was capable of fantastic things;' 'they disparaged everything Rosenblatt did, and most of what ONR did in supporting him;' 'a pack of happy bloodhounds;' 'Minsky knocked the hell out of our perceptron business;' 'Minsky and his crew thought that Rosenblatt's work was a waste of time, and Minsky certainly thought that our work at SRI was a waste of time;' 'Minsky and Papert set out to kill the perceptron, it was a sort of social service they could perform for the AI community;' 'there was some hostility;' 'we became involved with a somewhat therapeutic compulsion;' 'a misconception that would threaten to haunt artificial intelligence;' 'the mystique surrounding such machines.' These rhetorical expressions show the extent (the heat) of the perceptron controversy beyond doubt.

In the early 1960s, when the heat of the perceptron controversy was at its peak, Minsky and Papert decided to intervene decisively in it. They decided to 're-enact' (Latour, 1987) Rosenblatt's perceptron results. Minsky and Papert's involvement in the perceptron controversy at this point belongs to what Bruno Latour called the 'third way' of reading a scientific text (or some scientific evidence or results). Latour (1987, pp. 60-61) distinguished three ways in which a scientific paper can be read. 'Giving up' is the most usual one (this happens 90% of the time, according to Latour's informal estimates): people just do not read it. The second one is 'going along' (quite rare, about 9% of the time): the reader believes the author's claim and uses it (refers to it, and by doing so he or she helps transform the claim into a fact). The third way of reading a paper (an extremely rare and costly one) is 're-enacting' everything the author went through (and then at least one flaw can always be found even in the best written scientific text). Minsky and Papert embarked on a reading of Rosenblatt's perceptron papers of this sort. They re-enacted many of Rosenblatt's results — and they not only found

one flaw, but many (or so they claimed). It was costly for them: it took a lot of years and research effort. But, very importantly, as I will show in detail in sections 3.3 and 3.4, it became even costlier for neural network researchers to challenge their results. This move by Minsky and Papert can be seen, using Latour's terms, as one from weaker to stronger rhetoric.

The length (cost in terms of time) of this 're-enacting' process explains some questions related to the time frame of the perceptron controversy. The result of this re-enacting process — Minsky and Papert's 'Perceptrons' book — was not published until 1969. By then the crisis of early neural network research was well under way. Minsky and Papert's (1969) study was, as it were, the final 'push' for the definite closure of the controversy. Nonetheless, it is very important not to forget that many of Minsky's arguments against the perceptron had become well known since the late 1950s, and that he had actively defended them in the 'quarrels' that he had had with Rosenblatt in the late 1950s and early 1960s. By the mid-1960s Minsky and Papert's arguments against the perceptron (and neural networks in general) had had a significant effect on the crisis of early neural network research (this crisis is studied in section 3.2).

Minsky and Paper had worked (separately) on neural networks long before they embarked on their common 'Perceptrons' project in the early 1960s. Minsky's early work on neural networks was discussed in section 2.1; for Papert's involvement, see Papert (1988, p. 11). When Papert arrived at MIT in 1963, Minsky and Papert decided to start their project of a theoretically elaborated account of the limitations of perceptrons (what I earlier called the 're-enacting' of Rosenblatt's perceptron). This project proved to be costlier than expected, and Minsky and Papert's book did not come out until 1969. Minsky and Papert delayed publication until they had given a considerably elaborated mathematical form to many of their points. McCorduck (1979, p. 89) reported on this delay in her historical study of AI:

> "After working on the problem of Perceptrons for some three years, and coming to understand them at least partially, and proving some theorems about them, Minsky and Papert laid out their book. In the process of writing, loose ends appeared, and the two scientists kept working, tying up the loose ends and delaying publication."

In their new epilogue to their 1969 book, Minsky and Papert (1988, pp. 249-250) commented on one of the reasons for the delay.

> "It took us many months of work to capture in a formal proof our strong intuition that perceptrons were unable to represent that predicate [the connectedness predicate; see section 3.3]."

Seymour Papert (1988, p. 11) recently spoke of 'years of struggle.'

> "Minsky and I both knew perceptrons extremely well. We had worked on them for many years before our joint project of understanding their limits was conceived . . . Yet when we challenged ourselves to prove our intuitions it sometimes took years of struggle to pin one down — to prove it true or to discover that it was seriously flawed. I was left with a deep respect for the extraordinary difficulty of being sure of what a computational system can or cannot do."

In the following section (3.2) I study the crisis of early neural network research. It will be shown there that that crisis had reached considerable proportions by the mid-1960s. Later, in section 3.3, I will study Minsky and Papert's (1969) criticism of neural network in detail, but it is important to remember that many of the points of that criticism had had an important effect on the evolution of neural network research by the mid-1960s, much before the 1969 book was published.(this was confirmed by the early neural network researchers whom I could interview; see appendix 2).

In this section I have studied some of the rhetorical tactics used in the perceptron controversy by researchers both in favour and against neural network research. I have also discussed the beginning of Minsky and Papert's project aiming at showing decisively the limitations of the perceptron.

3.2 The crisis of early neural networks

In this section I discuss the problems that early neural network researchers were having with their machines and systems. I look at Rosenblatt's perceptron first, and I show that Rosenblatt was aware of the limitations of his single-layer machine. During the 1960s Rosenblatt tried to solve the problems of his perceptron machine, and kept working in perceptrons. Researchers belonging to Widrow's group and to the SRI neural network group were having increasing difficulties with their machines too. One problem which was studied by researchers from the three mentioned neural network projects was that of training multilayer networks. Towards the mid-1960s, both the Madaline and the Minos projects were in a situation of deepening crisis. I look at Widrow's project first, and then I examine the late developments of the SRI neural network project. Widrow and his colleagues started to apply their neural network techniques to areas outside neural network research (e.g. adaptive filters and adaptive antennas) successfully, and moved away from the neural network area. The SRI researchers also abandoned neural network research, and started to work in a robotics project within the symbolic approach to AI.

It was seen in section 3.1. that, in the 'heat' of the perceptron controversy, Rosenblatt was accused by his opponents of 'irritating' a lot of people with his 'exaggerated' claims about the capabilities of the perceptron. This was part of the rhetoric used by Rosenblatt's opponents in the perceptron controversy. It was seen in the previous section that rhetorical tactics were used heavily by both sides of the controversy. Nevertheless, as a result of the 'balance of power' established by the closure of the perceptron controversy (to be analysed in section 3.4), some elements of the rhetoric used by Rosenblatt's opponents (such as the mentioned one that 'Rosenblatt made exaggerated claims') became a part of the 'popular history' of neural network research

(an example of this was shown in the previous section). In this section I show that Rosenblatt and his colleagues were aware of the limitations of the perceptron, and often acknowledged them openly in the papers they published in the late 1950s and early 1960s.

One of the limitations most frequently acknowledged by Rosenblatt was the lack of capacity of the perceptron to detect similarities between figures in its retina. The reason for this was, as Rosenblatt (1958a, p. 96; 1962a, pp. 67-70; 1962b, pp. 390-391) openly admitted, that the perceptron did not classify objects according to their geometrical similarity. Instead, the classifications done by the perceptron were based on the amount of overlap or intersection between objects in its input retina. If the amount of overlap between the retinal areas occupied by two objects was big enough, then the perceptron could classify them under the same category.

For instance, the machine could, under the right circumstances, recognise the difference between two different kinds of stimuli (e.g. triangles and squares). But unfortunately 'under the right circumstances' here means, as Rosenblatt acknowledged, that two stimuli (presented one after another) had to occupy nearly the same area of the retina in order to be classified as similar. This means that inputs A and B in figure 3-1 would not be classified as belonging to the same category (square) by an elementary perceptron.

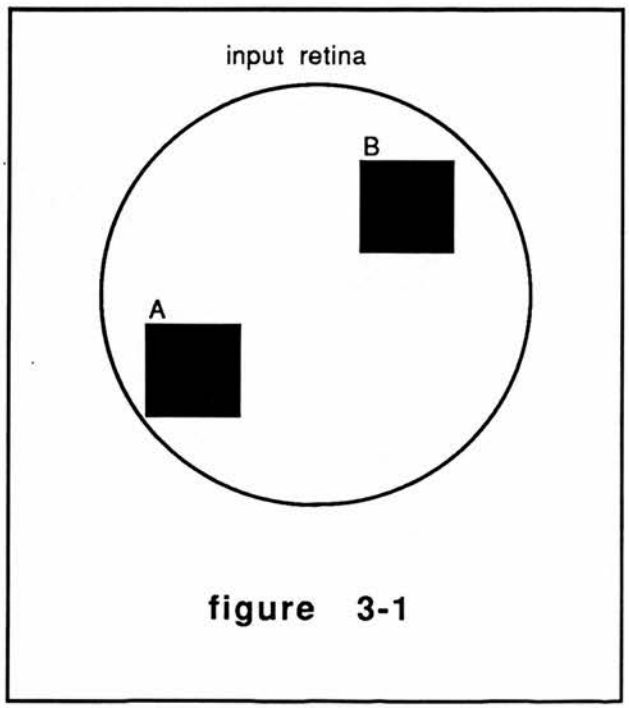


figure 3-1

Rosenblatt used the name 'weak generalisation' to refer to this type of overlap-based recognition, as opposed to 'pure generalisation,' which the elementary perceptron was *not* capable of:

> ". . . A pure generalization problem is one in which the . . . perceptron is required to transfer a selective response from one stimulus (say, a square on the left side of the retina) to a 'similar' stimulus which activates none of the same sensory points (a square on the right side of the retina) . . . The simplest of perceptrons [the single-layer perceptrons] . . . have no capability for pure generalization, but can be shown to perform quite respectably in discrimination experiments, particularly if the test stimulus is nearly identical to one of the patterns previously experienced."(Rosenblatt, 1962a, pp. 68-69)

The perceptron had other, equally worrying problems and limitations. Rosenblatt (1962a, pp. 306-310) summarised them. One of the problems, the issue of preprocessing (i.e. distinguishing the components of an image and the relationships between them), was related to the above mentioned question of the similarity criterion used by the perceptron. The lack of an

adequate preprocessing system meant that a set of association units had to be dedicated to the recognition of each possible object, and this created an excessively large layer of association units in the perceptron.

> ". . . The excessive size of the perceptrons necessary to deal with complex environmental situations is due largely to the necessity of having a characteristic set of association units representing every possible sensory field or sequence in its entirety. A preliminary coding of the field in terms of its parts and relations would greatly reduce the size of the system required to describe a given universe of situations." (Rosenblatt, 1962a, p. 306)

Other problems were excessive learning time, excessive dependence on external evaluation (supervision), and lack of ability to separate essential parts in a complex environment. Rosenblatt (1962, pp. 309-310) included the 'figure-ground' or 'connectedness' problem in this last point. This problem, the issue recognising a figure as distinct from its background, was one of the most important issues of Minsky and Papert's (1969) critical study of single-layer perceptrons (see section 3.3).

In dealing with the above mentioned problem of recognising similar objects appearing in different positions in the perceptron's retina (see figure 2-9), Rosenblatt studied systems with two layers of association units, and also systems with connections among the units of the same layer ('cross-coupled' perceptrons).[49] Rosenblatt (1962a, p. 576) claimed that the perceptron's generalisation capability improved considerably with these changes. In a perceptron with two layers of association units ('four-layer' perceptrons in Rosenblatt's terms), the units of the first association layer which responded to similar features in different positions would all activate the same unit in the second association layer, and in this way a

---

[49] In Rosenblatt's terminology perceptrons with two layers of association units were 'four layer' or 'multilayer.' However, today the term multilayer is used to refer to systems with more than one layer of *adjustable* connections. In other words, researchers use now the term 'layer' to refer to layers of modifiable connections, not to layers of units or to layers of fixed connections.

feature in different positions could be recognised as the same (von der Malsburg, 1986, pp. 245-246). Rosenblatt carried out some research on 'four-layer' perceptrons with one or more layers of modifiable connections. Nevertheless, Rosenblatt openly admitted that very important problems concerning 'four-layer' and 'cross-coupled' perceptron systems remained to be solved.

Rosenblatt (1962a, pp. 577-579) wrote a list of fifteen problems that perceptrons (more complex perceptrons included) had, some of which are reproduced in the quotation below:

> "A number of perceptrons analyzed in the preceding chapters have been analyzed in a purely formal way, yielding equations which are not readily translated into numbers. This is particularly true in the case of the four-layer and cross-coupled systems, where the generality of the equations is reflected in the obscurity of their implications . . . Those problems which appear to be foremost at this time include the following: (1) Theoretical learning curves for the error correction procedure . . . (2) Determination of the probability that a solution exists for a given problem . . . (3) The development of optimum codes for the representation of complex environments in perceptrons with multiple response units. (4) Development of an efficient reinforcement scheme for preterminal connections . . . (7) Theoretical analysis of convergence-time and curves for adaptive four-layer and cross-coupled perceptrons . . . (12) Effect of spatial constraints in cross-coupled systems (e.g., limiting interconnections to pairs of association units with adjacent retinal fields). (13) Studies of possible figure-segregation (figure-ground) mechanisms. (14) Studies of abstract concept formation, and the recognition of topological or metrical relations . . ."

(Here Rosenblatt uses the term 'terminal' to refer to the connections between the second association layer and the response units, and 'preterminal' to refer to the previous layers of connections.) This quotation shows beyond doubt that Rosenblatt was well aware of the considerable difficulties faced by early neural network researchers (and only a few of the

problems in his list are shown in the quotation). In points (4) and (7) above the difficulties of training multilayer networks, and in particular the lack of an effective algorithm for doing so, are clearly stated. Rosenblatt recognised that issues (4) and (7) were 'theoretical' (the word 'fundamental' would perhaps be more adequate here), i.e. that they could not be solved simply by carrying out more powerful simulations or by building more advanced machines:

> "In the case of problem (4) . . . simulation studies seem to be indicated for preliminary exploration, although it is hoped that some theoretical formulations may ultimately be achieved . . . The seventh question again is a theoretical one, although preliminary results obtained from simulation programs should prove enlightening." (Rosenblatt, 1962a, 579-580)

On the other hand, in point (13) of the above mentioned list of problems Rosenblatt insists once again on the figure-ground or 'connectedness' problem, which was one of the problems best analysed by Minsky and Papert (1969) in their critical study of perceptrons. It is important to emphasise that Rosenblatt's most pessimistic comments were for problems (13), i.e. the figure-ground problem, and (14), i.e. the recognition of topological relationships and abstract concepts:

> "These two problems [(13) and (14)] . . . represent the most baffling impediments to the advance of perceptron theory in the direction of abstract thinking and concept formation. The previous questions [from the first to the twelfth] are all in the nature of 'mopping-up' operations in areas where some degree of performance is known to be possible . . . [However,] the problems of figure-ground separation (or recognition of unity) and topological relation recognition represent new territory, against which few inroads have been made." (Rosenblatt, 1962a, pp. 580-581)

Rosenblatt (1958a, pp. 110-111; 1962b, pp. 390-391) and his colleague David Block (1962, p. 149) insisted often on the problems found in recognising topological and temporal relationships with the perceptron. These included predicates like

'the object to the left of the square,' 'the object which appeared before the circle,' 'the square is inside the circle,' and 'the dog is in front of the tree.' Rosenblatt openly acknowledged (see quotation above) that progress toward solving these problems, as well as the figure-ground problem, had been almost insignificant.

In 'Steps toward artificial intelligence' (one of the most important papers in early AI) Minsky insisted on the importance of both the figure-ground problem and the problem of recognising similar objects placed in different positions in the retina of the perceptron.

> "These nets [perceptrons], with their simple, randomly generated connections can probably never achieve recognition of such patterns as 'the class of figures having two separated parts,' and they cannot even achieve the effect of template recognition without size and position normalization (unless sample figures have been presented previously in essentially all sizes and positions). For the chances are extremely small of finding, by random methods, enough properties usefully correlated with patterns appreciably more abstract than those of the prototype-derived kind. And these networks can really only separate out (by weighting) information in the individual input properties: they cannot extract further information present in nonadditive form. The 'perceptron' class of machines have facilities neither for obtaining better-than-chance properties nor for assembling better-than-additive combinations of those it gets from random construction." (Minsky, 1961, p. 15)

Minsky's criticism of perceptrons will be studied in detail in section 3.3. The importance of (symbolic AI) machine vision research on the crisis of early neural network research will be studied later in this section, when I analyse the crisis of the Minos neural network project at SRI.

The problem of learning in perceptrons with more than one layer of adjustable connections (multilayer perceptrons in today's terms) was seen by Rosenblatt and colleagues as one of the most important difficulties for making further progress in perceptron

research. Rosenblatt had shown that, if a single-layer perceptron was able to embody a classification task, then it was capable of learning it after a finite (i.e. non-infinite) number of repetitions of the presentation of input/adjustment of weights cycle (see section 2.3). The problem was that there were some classification tasks that the single-layer perceptron could not realise. This problem was often shown in the simplest possible 'neural network:' a device with two input units and one output unit (and the corresponding two modifiable connections). It is easily shown (I will come back to this in section 3.3) that such a network can realise the 'and' function. In other words, values for the parameters of the system (connection weights and the threshold of the output unit) can be found which embody the 'and' function. The 'and' function is 1 only when both inputs are 1, and it is 0 otherwise. So the device of figure 3-2 below will only fire (i.e. will produce output 1) when both input units are activated (1, 1). Only in that case is the value of the threshold (1.5) exceeded.

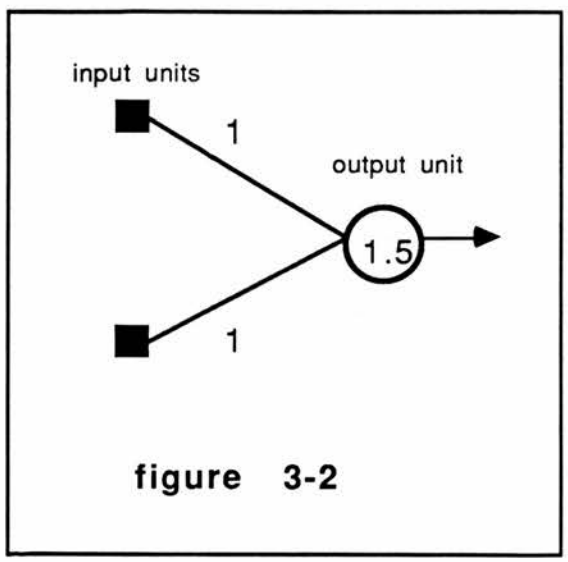


figure 3-2

Other functions such as the 'exclusive-or,' cannot be embodied by the system of figure 3-2. The output of the 'exclusive-or' function is 0 for inputs (1, 1) and (0, 0), and 1 for inputs (0, 1) and (1, 0). The network's structure cannot embody that function (and therefore the network can never learn such a classification). The device of figure 3-2 can only realise linearly separable functions, and exclusive-or is not linearly separable.

functions, and exclusive-or is not linearly separable. This issue will be discussed further in section 3.3.

What is important here is that early neural network researchers were aware of the problems of single-layer perceptrons long before they were studied in depth by Minsky and Papert. For instance, in J. K. Hawkins' (1961) review about 'self-organising systems' the issue of training multilayer networks figured prominently among the problems of single-layer neural networks. This review appeared alongside Minsky's (1961) above mentioned 'Steps toward artificial intelligence' paper in the same volume (vol. 49) of the 'Proceedings of the Institute of Radio Engineers' (IRE). Hawkins discussed the limitations of single-layer neural networks using the exclusive-or function as an example. It was well known then that some classifications which were not linearly separable could be realised with perceptrons with intermediate units ('hidden' units in today's terms). In the case of the system of figure 3-2 above, a hidden unit would send strong inhibitory activation when both inputs are 1 (this will be shown in section 3.3). The problem was that no weight modification rule had been developed which guarantied results comparable to those obtained with the perceptron learning algorithm or Widrow and Hoff's LMS algorithm.

> "For example, the AND [function] . . . can be realized with the [single layer] linear-logic circuit . . . while the exclusive-or [functions] . . . require a cascade linear logic arrangement ['hidden' units] . . . [The limitations of single layer networks] are extremely severe . . . since the percentage of realizable logical functions becomes vanishingly small as the number of input variables increases. The chances of obtaining an arbitrary specified response are correspondingly reduced. More sophisticated approaches must therefore be undertaken. A number of alternatives are possible . . . The most attractive appears to be multiple-layer logical circuit arrangements, since it is known that any function can thereby be realized . . . However, no general criteria on the basis of which intermediate logical layers can be taught functions required for over-all network realization of the desired

input-output relationship have been discovered." (Hawkins, 1961, pp. 45-47)

Later in this section it is shown that the issue of learning in multilayer systems figured prominently in the agenda of early neural network researchers other than Rosenblatt too. In later sections it will be seen that developments around this question have been of great importance throughout the history of neural network research.

So far in this section it has been shown that Rosenblatt and his colleagues had considerable problems with their perceptron in the 1960s. Nevertheless, Rosenblatt did not give up, and continued doing research on perceptrons, trying to develop solutions for the limitations of his first machine. Later developments in Rosenblatt's project will be discussed in section 3.4.

Now I will examine the crisis of Widrow's group's neural network project, and later I will look at the late stages of the neural network project of Rosen and colleagues at SRI.

Widrow and his colleagues were aware that their most powerful neural network machine, the single-layer Madaline (with one layer of adaptive connections, see figure 2-9), had important limitations. For Widrow and colleagues, learning in multilayer systems was a very important problem. They were aware that there were important classifications that the (single-layer) Madaline could not realise. It was known that machines with two layers of modifiable connections have much greater classification power. It is important to note that Widrow and his students studied multilayer madalines, and investigated learning procedures for them. These were madalines in which the second layer of connections was also adaptable. Figure 3-3 below shows one type of multilayer architecture studied by Widrow and his colleagues.

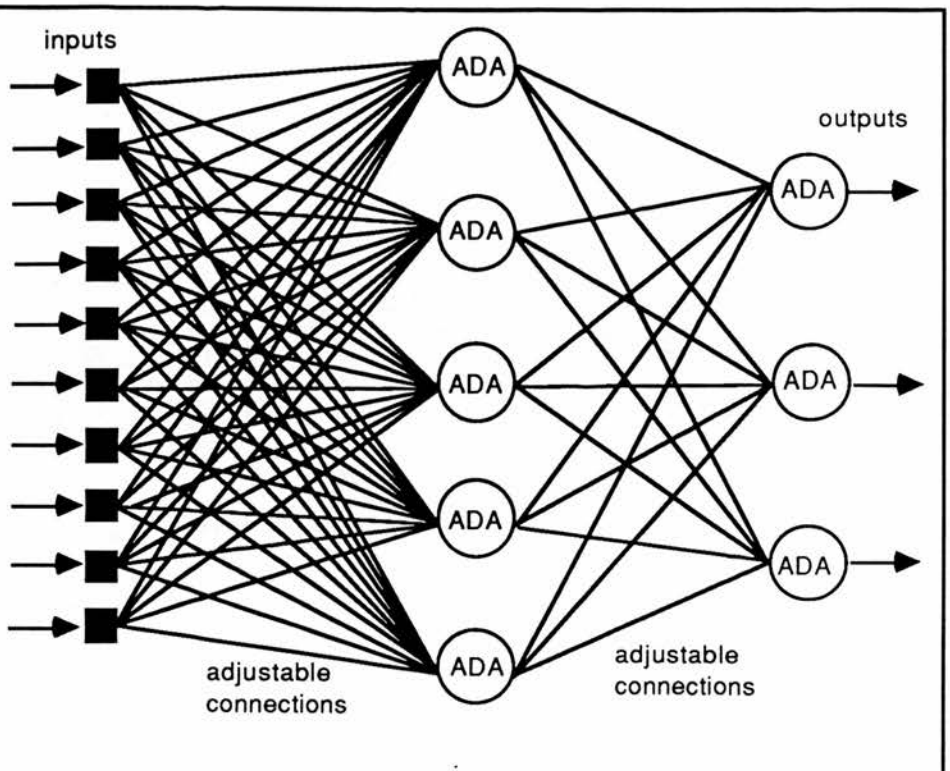


figure 3-3: madaline with two layers of modifiable connections

The system represented in figure 3-3 above is capable of classifying inputs into eight categories (it has three binary output units). In some of his experiments Widrow used a 4x4 square retina (i.e. 16 input units which could represent, e.g., letters). He also studied systems with bigger input 'retinas.'. The objective of Widrow's (1962, p. 455) experiments with the multilayer network of figure 3-3 was:

> ". . . To teach the system to classify . . . patterns [belonging to the 8 categories] correctly by showing it only a very small randomly selected fraction of the total number of possible input patterns. If the first layer could be trained to produce a set of output signals which are close to being independent of rotation, translation, size, and noise, the second layer could be trained to produce the specific desired responses. To do all this, some of the adalines

shown in figure . . . [3-3] might have to be madalines, or a third (or more) adaptive layer might have to be added."

Widrow (1962, p. 456) described a procedure for training a system like the one on figure 3-3 which 'had been found by experiment to work well:'

> "The first layer is adapted first in an attempt to get the second layer outputs to agree with the desired outputs. The first layer neurons to be adapted are chosen to minimize the number of adaptions . . . If no combination of adaptions of the first-layer neurons produces the desired outputs, the second layer neurons should then be adapted to yield the desired outputs. This procedure has the tendency to force the first layer neurons to produce independent responses which are insensitive to rotation. All adaptions are minimum mean-square error. Again, responsibility is assigned to the neuron or neuron that can most easily assume it."

The expression 'had been found by experiment to work well' above means that results could not be guarantied in the same sense as the LMS algorithm guarantied results for single-layer machines.

Widrow (1962, p. 456) reported some 'successful' experiments on 'wide varieties of specific responses' with a small scale version of the system of figure 3-3, namely a system with three adalines in the first layer of units and one in the second (and therefore two layers of modifiable connections).

Widrow and colleagues dedicated considerable research efforts to multilayer machines.

> "The above [adaption] procedure and many variants upon it are currently being tested with larger networks, for the purpose of studying memory capacity, learning rates, and relationships between structural configuration, training procedure, and nature of specific responses and generalizations that can be trained in." (Widrow, 1962, p. 456)

Many experiments were done, but Widrow and colleagues were not able to develop a learning algorithm for multilayer networks comparable to the one they had developed for single-layer networks (the LMS algorithm; see section 2.3). In part as a result of that, they started to concentrate their attention on engineering applications of adalines and the LMS algorithm outside neural network research. I asked Widrow about his group's research in training multilayer networks in the early 1960s. This was his response:

> "We tried to adapt layered neural nets, and never succeeded, except, we were able to adapt a two layer network, the madaline, where the first layer was adaptive but the second layer was fixed. By knowing the nature of the second layer we were able to make rules for adapting the first layer. But if the second layer was completely free to do what it wanted, we didn't have any general rule for adapting the first layer . . . We tried to adapt multilayer networks, we were trying to make that breakthrough, and tried, and tried, and tried, but never succeeded. I couldn't imagine anything to do it, any way to do it. I think that, if anything, for lack of success on that I switched and stopped working on neural networks." (Widrow, interview)

Widrow recently made a similar point in his '30 years of adaptive neural networks' paper:

> "After devising their Madaline rule, Widrow and his students developed uses for the Adaline and Madaline . . . Work then switched to adaptive filtering and adaptive signal processing, after attempts to develop learning rules for networks with multiple adaptive layers were unsuccessful." (Widrow & Lehr, 1990, p. 1415)

The crisis of Widrow and colleagues' neural network research project reached its peak towards the mid-1960s. By that time, further progress in neural networks started to look increasingly difficult to them. At about the same time, applications of their ideas (mainly of the adaline and the LMS algorithm) in areas like adaptive filtering and adaptive signal processing (see Widrow & Stearns, 1985) started to be more successful (Widrow,

interview). Widrow (interview) recalled the departure of his colleague Marcian Hoff from Stanford, which happened around 1965, as a significant turning point in the research interests of his group. The work carried out by Hoff after he left Widrow's laboratory at Stanford University was quite remarkable. He was credited with the invention of the microprocessor (he got IEEE fellow membership 'for the invention of the microprocessor').

> "I think that he [Hoff] left to join Intel [Silicon Valley, California] in about 1965 or 1966. When he went to 'Intel' he had an interesting idea, that he could use the electronics that they had to build an entire computer on one chip. So he built the first microprocessor, and he is regarded all over the world as the inventor of the microprocessor. He has become a fellow of the IEEE, and his citation reads 'for the invention of the microprocessor.' He has had a very nice career. So that's what happened to the other half of the LMS algorithm." (Widrow, interview)

One first successful application of Widrow's adaline and LMS algorithm (fist developed in neural network research) outside neural networks was in adaptive antennas (see Widrow et al., 1967).

> "At the time that Hoff left, about 1965 or 1966, we had already had lots of troubles with neural nets. My enthusiasm had dropped. But we were beginning to have successful adaptive filters, in other words, finding good applications. We were using the LMS algorithm to adapt both neural nets and adaptive filters. I had some very good success with adaptive antennas. We were making antennas that had the capability of receiving a signal from any direction that you wished and, if anyone tried to jam that antenna, the antenna would automatically reduce its sensitivity in the direction of interference. It learned all by itself, using the LMS algorithm. It's just taking an antenna . . . and connecting a neural net to it, but a neural net without the quantisers [thresholds]. It was just a single neuron without non–linearity. It worked unequivocally, you can prove it mathematically, we were all delighted, we were very happy with it. So you are happy with something, and another thing [neural networks] is frustrating and it can't overcome certain problems. So

> guess which direction you are going? So we stopped, basically stopped on neural nets, and began on adaptive antennas very strongly." (Widrow, interview)

But adaptive antennas were not the only successful application for adalines and the LMS algorithm developed by Widrow and others. In the second half of the 1960s, R. W. Lucky and his team at Bell Laboratories (Lucky, 1965; Lucky et al., 1968) applied adaptive filters to telephone systems (Widrow & Lehr 1990, pp. 1415-1416). One of the application areas was adaptive equalisation in high-speed modems. By using adaptive filters in high-speed digital data transmission, the amount of data sent through the same telephone channel can be increased four times without loss of reliability (Widrow, interview). For that, a modem which has an adaptive filter on it has to be installed in the receiving telephone.

Another application of Widrow's adaline and LMS in the area of telephone systems, also developed at Bell Laboratories, was an echo cancelling device for long-distance telephone and satellite circuits. This echo canceller had an adaptive filter capable of doing LMS weight adjustment. By using it, telephone line echo is cancelled, and it is also possible to have communication in both directions at the same time (other echo cancellers prevent communication in two directions at the same time) (Widrow, interview). This device is specially useful for long-distance and satellite telephone communication because of the longer time delays produced in these cases. Furthermore, this echo canceller can also be installed in modems, so that echo can be avoided in high-speed digital data transmission by telephone too.

> "So modems also have echo cancellers, because you can't have echo — even the slightest trace of echo — when you are transmitting digital data at high speed. So you can have an adaptive filter in the modem, for echo cancelling, and another one for equalisation, and they must adapt to that particular phone line, because every phone line is different, so the receiving filter must adapt to that line, you can't use a fixed filter." (Widrow, interview)

Because of all these developments, Bernard Widrow was awarded the IEEE 'Alexander Graham Bell' prize in 1986 for 'exceptional contributions to the advancement of telecommunications' (Widrow, interview).

It is important to note that, after the crisis of early neural network research, both the LMS algorithm and the adaline (as an adaptive filter), first developed in the context of neural network research, were rather successful in other application areas leading, as Widrow and Lehr (1990, pp. 14-16) pointed out, to "major commercial applications." Widrow (interview) described the applications of the adaline and the LMS learning algorithm in areas other than neural networks in the following terms:

> "I didn't invent the echo canceller, I didn't invent the adaptive equaliser, Lucky invented the adaptive equaliser. But what Hoff and I invented is the LMS algorithm. I had been using it for years on adaptive filters. You see, no person does all this, it's a combination of contributions that accumulate together to make these things possible. Now this adaptive equaliser is so popular, that even the cheapest modems have it. It has a chip on it that does LMS, so it's got effectively one neuron inside the equaliser, it's a neural net with one neuron. LMS has been a successful algorithm, probably the most widely used one in the world of adaptive systems today, and now it's so old, from 1959, when we discovered it. The LMS algorithm was developed for neural nets, then used in adaptive filters, and now, 20 years later, it is used back in neural nets again."

In section 5.3 it will be shown how, more than twenty years after the crisis of early neural networks, Widrow was 'rediscovered' in the re-emergence of neural network research in the 1980s. After being rediscovered Widrow became a leading member of the new neural network community in the late 1980s.

So far in this section I have looked at the crisis of two early neural network projects, namely Rosenblatt's and Widrow's. Now I will examine the crisis of the neural network research project at Stanford Research Institute (SRI).

By the mid-1960s financial support for the SRI neural network project (see section 2.4) was running out and the limitations of the Minos machine were becoming increasingly apparent. Nils Nilsson (1965, ch. 6), from the SRI group, carried out a study of those limitations. The members of the SRI neural network group whom I interviewed in the United States (namely Rosen, Nilsson, Duda, and Hart) confirmed to me that by the mid-1960s the SRI group had started to switch its research interests from neural networks to (interestingly) the symbol-processing approach to AI.

> "When we stopped the neural net studies at SRI, research money was running out, and we began looking for new ideas. It was getting harder to do a little more each time, and it didn't look like it was worth that much." (Rosen, interview)

Nils Nilsson, another leading member of the SRI group, described the end of SRI's neural network research project in these terms:

> "About 1965 or 1966 we decided that we were more interested in the other artificial intelligence techniques. I still thought that neural networks would have some use at some future time, but I thought that we had reached pretty much as far as we could go." (Nilsson, interview)

The SRI researchers were aware of the problems and limitations of single-layer neural networks (Nilsson, 1965), and the prospects for more complex machines did not look very encouraging to them. The important problems which had to be solved — and the SRI researchers could not solve adequately — in order to design and study more complex neural network systems included (yet again) the issue of training multilayer networks. The SRI researchers were aware that more classification power could come from making the connections on the second layer of their Minos machine adjustable, but they could not develop adequate training techniques for that. This was openly admitted by Nilsson (1965) and confirmed to me by Nilsson himself (interview) and by Rosen (interview).

> "In general, layered machines can be trained by varying the weights associated with *each* TLU [threshold logic unit or 'neuron'] in the network. There do not exist, however, efficient adjustment rules for such thorough training of a layered machine . . . The committee machine [Minos or the madaline] can be generalized by allowing the committee [majority logic] TLUs to have different voting strengths . . . The possibility of such variants of the committee machine increases its classifying power but, unfortunately, no efficient training procedures are known which simultaneously locate the weight vectors and adjust their voting strengths." (Nilsson, 1965, pp. 97-99)
>
> "I got very interested for a while in the problem of training more than one layer of weights, and was not able to make very much progress on that problem." (Nilsson, interview)
>
> "Our group never solved the problem of training more than one layer of weights in an automatic fashion. We never solved that problem. That was most critical. Everybody was aware of that problem." (Rosen, interview)

Of course, this does not mean that training multilayer systems was the only problem faced by SRI neural network researchers. In section 2.4 some other problems were commented, and much of what has been said in this section about the problems found by Rosenblatt and Widrow applies to the Minos project researchers too.

One aspect of the crisis of the Minos project which is of particular interest here is the change of research interest at the SRI group from neural networks to robotics and machine vision within the broad symbol-processing 'umbrella.'

Work on machine vision within the symbol-processing approach, mainly that of Larry Roberts (1963) of MIT Lincoln Lab., had a significant impact on the SRI researchers, especially on those more oriented towards pattern recognition like Richard Duda and Peter Hart (Duda, interview; Hart, interview). Roberts designed a system for transforming digitised pictures into line drawings, which were then compared with stored information about

geometric properties of objects belonging to a simple block world (Fleck, 1978, pp. 94-101). Emphasis on main stream AI machine vision research was on scene analysis in terms of lines, edges, vertices, relative brightness, and linguistic descriptions, rather than on neural network-like pattern classification. For this, the computer system needed an internal representation or model of its surrounding block world (its expectations or knowledge about that world).

The SRI researchers were impressed by this AI research on machine vision at about the same time as they were having considerable trouble with their Minos neural network machine. Machine vision was a part of the symbolic AI approach, but it had its own peculiarities. It was closer to the SRI researchers' interests in perception and vision than symbolic AI models of 'higher' cognitive processes, and also it was pretty much connected to robotics research, a continuing interest of Charles Rosen and his colleagues.

By the mid-1960s the symbol-processing approach to AI had emerged with considerable success (see section 2.5). Broadly speaking, symbolic AI researchers were using digital computer programs to model aspects of intelligent human behaviour. Within this context, as funding for the Minos project was running out and problems like training multilayer systems could not be adequately solved, Rosen and his colleagues decided to abandon the neural network approach and to start a robotics project within the symbol-processing perspective. They hired Bertram Raphael, a former student of Marvin Minsky, to teach them LISP and help them in their new project: a robot named 'Shakey.' LISP (list processing language), developed by John McCarthy in 1960, had become the most widely used programming language in symbolic AI.

> "We hired Bert Raphael from MIT who taught us LISP. We were interested in learning LISP programming, and that started to be more interesting than neural networks." (Nilsson, interview)

The fact that the SRI researchers had not had any experience with LISP programming by the mid-1960s shows the 'barrier' that separated neural network researchers and symbolic AI researchers at that time. Pamela McCorduck (1979, p. 231) reported on this in her historical study of AI.

> "Raphael . . . had been hired by SRI for this project [the Shakey robot] as the only one who knew LISP and who had had experience with the LISP language and large computers."

The arrival of Raphael, a student of Minsky (the 'spokesman' or 'symbolic leader' of the anti-neural network position in the perceptron controversy), was a clear sign of the change at SRI. According to Peter Hart (interview) who joined the SRI group in 1966, by that time the point of view was much more symbolic AI than neural networks:

> "By the time I joined the group, 1966, the point of view was much more computational architecture than it was networks of devices. At that time, in 1966, we started the famous 'Shakey' robot program, and there the point of view was strictly what kind of computer program, what kind of representation do we need inside the computer to enable a robot to deal with various kinds of real world phenomena. By the late 1960s perceptrons, adalines, learning machines, by that time all that was pretty much over. By that time people thought that it was not the most promising approach. I think that the approach had intellectually run out of esteem."

Richard Duda confirmed the change in the research interests of the SRI group in the mid-1960s from neural networks to symbolic AI. He also pointed out that Minsky was aware of the neural network developments at SRI. (This is not new. Remember Rosen's complain that 'Minsky knocked the hell out of their perceptron business' in section 3.1. Minsky's role in the perceptron controversy will be further discussed in sections 3.3 and 3.4.)

> "There was a growing interest in [symbolic] artificial intelligence. By the time Raphael joined the group, the group became a [symbolic] artificial intelligence centre. Raphael was one of Marvin Minsky's students. There had been connections between Minsky and SRI before that too. Marvin was a consultant for us on a couple of occasions. He was certainly quite familiar with Minos." (Duda, interview)

It is likely that Minsky was quite influential in the decision of the SRI group of changing their research from neural networks to symbol-processing. Minsky told me that the change of research direction at SRI is a 'good example' of the crisis of early neural networks.

> "A good example, SRI, had given up perceptrons by that time [the mid-1960s]. They hired Raphael, one of my students. They started to use LISP, and they became one of the great centres of heuristic programming. They got the 'Shakey' robot, and things like that. By that time the perceptron project was really dead." (Minsky, interview)

The perceptron project was in crisis, but not completely dead. Rosenblatt and colleagues were still working in perceptrons (see section 3.4).

In this context, the SRI researchers decided to apply for funding for a robotics project. The idea was to build a mobile robot (which they later named 'Shakey'). Robotics was closer to their interest than other areas of symbolic AI research (e.g. game playing, theorem proving, or natural language). At the same time as the SRI researchers started their robot project, several other important AI robotics projects were started in the mid-1960s at MIT, Stanford University ('Pingle') and SRI, and later in Edinburgh University ('Freddy') (Fleck, 1978, pp. 114-121). The SRI group became a leading symbolic AI centre. Some of these robotics projects (MIT, Standford University, and Edinburgh University) were 'hand-eye' systems, but 'Shakey' was a mobile robot which could navigate in a room containing large blocks (obstructions). Shakey could also carry out simple tasks like taking a block from one room to another.

One of the most interesting aspects of these robotics projects of the early 1970s was the integration in the same system of interacting subsystems for vision, planning, and object manipulation. The SRI researchers combined perceptual, motor-control, problem solving, and knowledge-representation systems in their 'Shakey' robot (McCorduck, 1979, pp. 223-235). The quotations below by Nilsson and Raphael, two leading researchers of the SRI robotics group, show the importance in 'Shakey' of symbolic AI themes such as internal representation and models of the (block) world, problem solving, and plans.

> ". . .'So there's an interesting research area [Raphael speaking] that we made some progress on — how to build robust systems, and what kinds of monitoring are needed and how the system has to check whether it accomplishes what it tries to accomplish. We developed ways of using the TV camera and sensory feedback to monitor and update Shakey's own model of the world. We built various ideas of representing information in the robot's mind as in a computer. In a sense, the robot has a model of itself and of its environment.' . . ." (Raphael, as quoted in McCorduck, 1979, p. 232)

> ". . .'Those of us at SRI [Nilsson speaking] were . . . interested . . . in general problem-solving mechanisms for reasoning out the solutions to problems . . . We also concentrated . . . on the interaction between the plan that was developed by the problem-solving system and the execution of that plan.'. . ." (Nilsson, as quoted in McCorduck, 1979, p. 231)

It is interesting to note that, although Minsky encouraged this type of projects, he was in favour of giving more priority to higher level processes.

> ". . .'You might say [Minsky speaking] that making robots was a sort of hobby which I encouraged but didn't really concern myself with that much, and I always felt that studying the sensory and perceptual systems is not the best way of thinking about thinking . . . The things I was most concerned with were the theses like Slagle's and Bobrow's and Raphael's, and such people who were really

> working on the symbolic problem-solving things.' . . ." (Minsky, as quoted in McCorduck, 1979, p. 224)

Charles Rosen, from the SRI group, played an important role in getting funding for the 'Shakey' mobile robot from the Advanced Research Projects Agency (ARPA, today's DARPA) of the US Department of Defense. Rosen pointed out that it was difficult to 'sell' the project to the people who did not like perceptrons, and who knew that the SRI researchers had been working on perceptrons until then.

> "By the time when we stopped the neural network project at SRI, computers had become available, simulations were possible. I gathered together a group of about 15-25 people. We brainstormed, meeting once or twice a week, and asked: 'what project shall we select to get into the main fields of artificial intelligence? Starting with what we knew about — neural nets — and going from there to artificial intelligence.' It took about 3 to 4 months, with a lot of ideas being examined. We then decided to propose making a robot, a mobile robot. It took me and my colleagues one year and a half to two to sell that programme to ARPA. It was very difficult, we had to sell it to some of the people who didn't like perceptrons, and that was our background, but on the other hand we had a crew of very able people. Finally we got ARPA money, and for 6 or 7 years we built, I'd say, the first really smart robot in the world." (Rosen, interview)

For more information on 'Shakey,' see Nilsson and Raphael (1967), and Raphael (1976, pp. 275-282). There were some controversies surrounding how funding for the 'Shakey' project was given and — years later — cut.

> "Rosen recalls how they found someone in the defense department who was willing to support the research, though for what Rosen himself considered foolish reasons, namely, that somehow a robot could be developed that could go about surreptitiously gathering information — a mechanical spy" (McCorduck, 1979, p. 233).

For the controversy about the cut in ARPA funding for 'Shakey' in the 1970s, see (ibid., pp. 233-235). 'Shakey' became quite

popular outside the scientific community thanks to a controversial article published in the magazine 'Life' (Darrach, 1970), which was widely read (McCorduck, 1979, p. 235) (I made some comments about this article in section 3.1). The cut in robotics funding was a general phenomenon, and was related to Lighthill's (1973) negative report about robotics in Britain. Although important AI ideas (e.g. in machine vision, or others like Minsky's 'frames'), as well as hardware devices, had been developed in the robotics projects of the late 1960s and early 1970s, funding for robotics was significantly cut in the 1970s.

What is of interest here is that the change in the direction of research at SRI reflects clearly the situation of deepening crisis of neural networks and institutionalisation of symbolic AI. For the SRI group the 'Shakey' project meant the change from neural network research to symbolic AI, and in particular to machine vision and robotics. Unlike in Minos, in 'Shakey' symbolic representation and knowledge issues became increasingly important, and much attention was devoted to the combination of problem solving and reasoning mechanisms with sensory and vision processes.

In this section I have looked at some of the problems that early neural network researchers were having with their systems. One particularly worrying problem was that of training multilayer systems. Early neural network researchers were well aware of this problem but, by the mid-1960s, they had not developed adequate solutions for it. I have showed how, in the mid-1960s, the researchers belonging to both Widrow's group and the SRI group changed from neural networks to other areas of research. Widrow started to apply his techniques in adaptive signal processing with increasing success. Researchers at SRI changed from neural networks to symbol-processing AI, and developed a project in robotics. Rosenblatt and his colleagues kept working in neural networks, but they were increasingly isolated, as I will show in section 3.4.

3.3 Interpretative flexibility

The 'heating' of the perceptron controversy, studied in section 3.1, was only the beginning. Rhetoric soon went from weaker to stronger. In this section I study the results of Minsky and Papert's (1969) project of re-enacting Rosenblatt's perceptron. I show that their study had two important parts. One was about the limitations of the single-layer perceptron; the other was an intuitive judgement about learning in multilayer perceptrons. Minsky and Papert's (1969) study is usually taken as the 'proof' that perceptron research had so many problems that it was not worth pursuing. But this common view is the *result* of the closure of the perceptron controversy (section 3.4). Before then, things were not so clear at all. In this section I show that the two main parts of Minsky and Papert's attack on the perceptron were open to interpretative flexibility. As with rhetoric, interpretative flexibility here went from weaker to stronger. And not only 'in principle.' Neural network researchers took advantage *in practice* of that interpretative flexibility in order to launch a counter-attack. The problem was, as always, whether the counter-attack was strong enough (I will come to this in section 3.4). In this section I show that there was nothing intrinsically compelling in Minsky and Papert's study of the limitations of the perceptron.

At the end of section 3.1 it was said that in the early 1960s, as the perceptron controversy was increasing, Marvin Minsky and Seymour Papert of MIT AI laboratory decided to make a decisive move: they decided to 're-enact' (Latour, 1987) Frank Rosenblatt's perceptron results and show conclusively the flaws they contained. In other words, Minsky and Papert decided to intervene in the perceptron controversy using as strong and decisive a rhetoric as possible, so that the debate would be

settled once and for all. The move here was not from rhetoric (see section 3.1) to 'truth' or 'rationality,' but, as Latour puts it, from weaker rhetoric to stronger rhetoric. Minsky and Papert decided to mobilise as many (and as good) allies and 'actants' as possible in their favour, so that their position could not be contested this time round.

Latour (1987, p. 60) pointed out that, when such a re-enacting process is carried out, then at least one flaw can always be found even in the best written scientific test (this is another way of saying that every scientific result is, in principle, open to interpretative flexibility). After re-enacting the perceptron, Minsky and Papert claimed that they had found not one, but many. There were two main issues in Minsky and Papert's lengthy and costly (in terms of time and research effort) project. On the one hand they studied some important limitations of the single-layer perceptron; on the other, they formulated a challenge about perceptrons more complex than those studied by Rosenblatt (and by themselves). If these two issues were accepted, then there would be little room for neural network researchers to manoeuvre, and controversy would be closed against neural network research. In this section I will analyse these two issues in detail, and I will examine their importance in the perceptron controversy.

In this study of Minsky and Papert's (1969) criticism I will use two concepts which need some clarification, namely the concept of 'anomalous' problem and the concept of 'reverse salient'. The concept of 'anomalous' problem will be used in the discussion of the first issue mentioned above, namely the limitations of the (single-layer) perceptron. The concept of 'reverse salient' will be used in the discussion of the second issue: the challenge about more complex perceptrons.

Thomas Kuhn (1970) used the term 'anomaly' extensively in his historical studies of science. One definition of anomaly refers to results which do not fit within the accepted categories of a scientific theory (in this sense of the term, when anomalies pile

up and are important and persistent, their solution may require severe adjustments in the conceptual apparatus of a theory). Another definition of anomaly — more in line with my concerns in this section — is simply a puzzle which resists solution. Other definitions have been proposed. Larry Laudan (1977, p. 29) suggested that an empirical problem becomes an anomaly for a theory (or research tradition) when it has not been solved by that theory but has nonetheless been solved by an alternative theory in the same research domain:

> "Whenever an empirical problem, p, has been solved by any theory, then p thereafter constitutes an anomaly for every theory in the relevant domain which does not also solve p."

Laudan's definition is not useful because it ignores — as philosophers sometimes tend to do — the social character of problems and anomalous problems in science. Barry Barnes (1982, p. 100) was suggesting this when he said that:

> "What one scientist sees as an anomaly another sees as a puzzle for the same paradigm — even a successfully solved puzzle."

In this section I will use the term 'anomalous' instead of Latour's above mentioned 'flaws' to refer to puzzles or problems which were especially worrying for neural network researchers. They became 'anomalous' not because of their 'intrinsic' or 'necessary' character, but because early neural network researchers were not strong enough to resist Minsky and Papert's (1969) conclusions about them. Those puzzles became anomalous because they were important in the closure of the perceptron controversy. But that closure was a social process, and the 'anomalous' character of those problems was the result of such a process, and not its cause. There was nothing 'intrinsically' anomalous in the problems analysed by Minsky and Papert. Quite the opposite, before the perceptron controversy was closed, there were different interpretations of those problems, some of them favouring further neural network research.

I will concentrate on two puzzles studied by Minsky and Papert (1969): the 'parity' problem and the 'connectedness' problem. Figure 3-4 shows the single layer perceptron analysed by Minsky and Papert in their study.

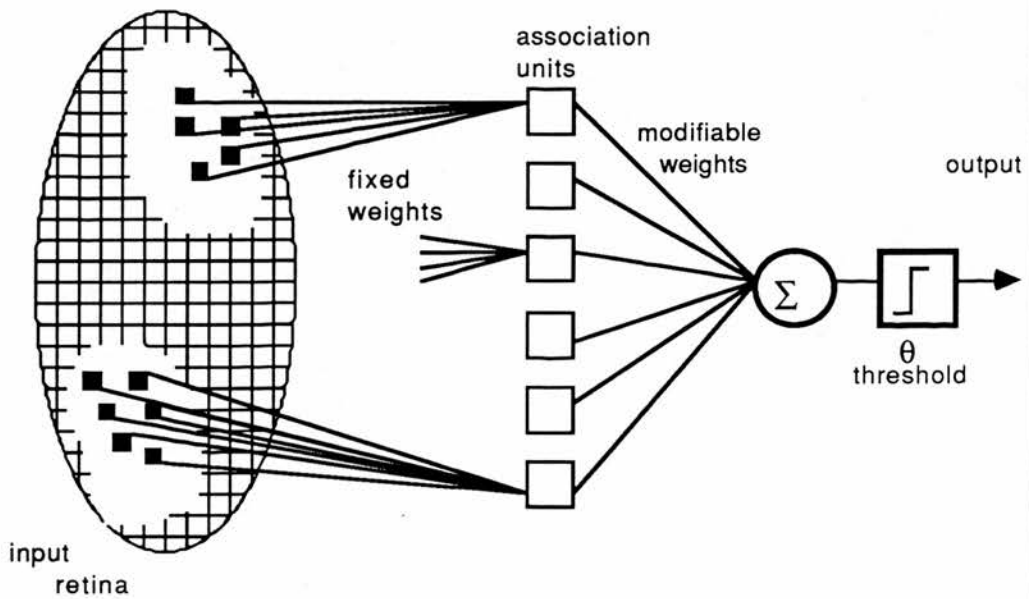


figure 3-4: perceptron studied by Minsky and Papert

Minsky and Papert introduced some important restrictions in the perceptron they studied. A very important restriction referred to the connections from the input units to the association units (the layer of fixed connections in the figure above). Minsky and Papert (1969, p. 5) pointed out that: "if we do not make restrictions, we do not get a theory." The restriction in the input-to-association connections was a consequence of Minsky and Papert's definition of computing in a perceptron. They defined the computation realised by a neural network system as a parallel combination of local information. Minsky and Papert thought that, for this computation to be interesting, it had to be simple in some meaningful sense. Minsky and Papert (1969, p. 9) defined this criterion of parallel combination of local information in the following way:

> ". . . The definition of *conjunctive localness* . The intention of the definition was to divide the computation of a predicate ψ into two stages. Stage I: The combination of many properties or features φ_α which are each easy to compute, either because each depends only on a small part of the input space R, or because they are very simple in some other interesting way. Stage II: A decision algorithm Ω that defines ψ by combining the results of the Stage I computations. For the division into two stages to be meaningful, this decision function must also be distinctively homogeneous, or easy to program, or easy to compute."

The computation realised by the output unit in figure 3-4 (stage 2 in the quotation above), a sum of incoming weighted activation in parallel plus a comparison with a threshold, satisfies Minsky and Papert's criterion. In the case of the association units (stage 1 in the quotation above), Minsky and Papert interpreted their 'simple combination of local information' criterion as implying that each association unit could not receive connections from many input units. This was quite consistent with Rosenblatt's ideas about input-to-association connections. In the Mark 1 machine there were 400 input units (20x20 retina), 512 association units and 8 output units. Each association unit was allowed to receive up to 40 incoming connections (10% of the input units) (Bernstein, 1981, p. 95). In the perceptrons studied by Rosenblatt (1962a) this number was usually smaller. Minsky and Papert's 'simple, parallel combination of local information' criterion meant, in the case of input-to-association connections, that each association unit could receive incoming connections only from a small part of the input retina. Minsky and Papert defined the 'order' of a perceptron as the maximum number of incoming connections received by an association unit. They also studied diameter-limited perceptrons, in which the restriction is the diameter of the circle (or set of input points) at which an association unit can 'look at.' The order of the perceptron of figure 3-4 is 6, because that is the maximum number of connections received by any association unit.

The implications of Minsky and Papert's criterion of conjunctive localness are better understood by analysing the particular problems that they studied. I will concentrate on two problems here, the 'parity' problem and the 'connectedness' problem.

The parity problem consists of saying whether the number on activated inputs in a perceptron retina (set of input units) like the one on figure 3-5 is odd or even. In figure 3-5 the number of inputs which are 'on' is odd (namely 13 activated inputs).

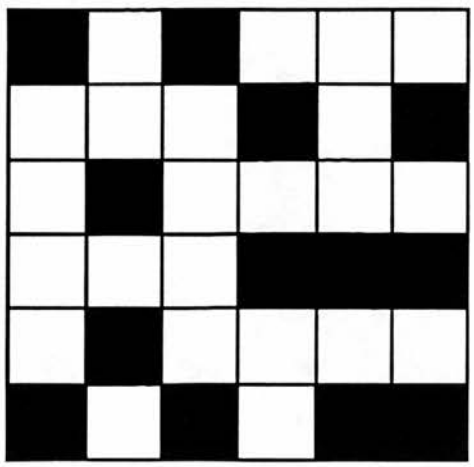


figure 3-5: odd number of activated inputs

The problem of parity is related to the 'exclusive-or' logic function. In a network with two input units and one output unit, computing parity is equivalent to computing exclusive-or. A system with two input units and one output unit can compute the 'and' function, as figure 3-6 shows, by giving each connection a value of 1, and the threshold a value of 1.5. The only case in which the output should be 1 is when both inputs are activated (1, 1), and in this case the value of the threshold would be exceeded (2>1.5). In all the other cases (i.e. inputs 10, 01, and

00) the summed activation value received by the output unit (1, 1, and 0 respectively) would not exceed the threshold value (1.5).

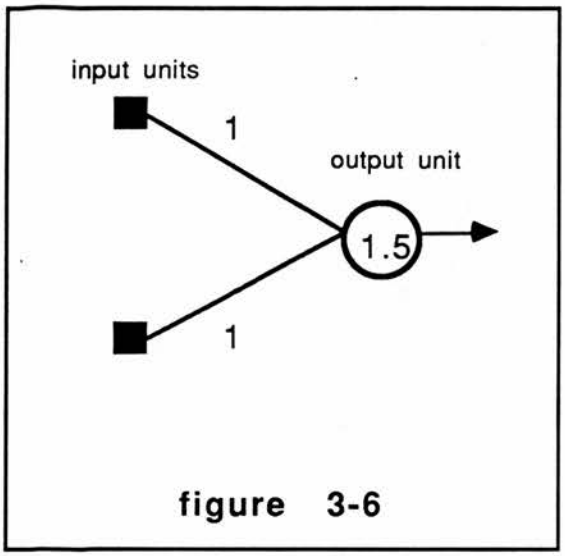


figure 3-6

The input space of the network of figure 3-6 can be represented as a two dimensional space (see figure 3-7). The computation realised by the output unit (where weights and threshold intervene) separates that input space into two regions, one corresponding to output value 1 (in bold in the figure below) and the other one to output value 0. The 'and' function is said to be linearly separable because a straight line that separates the two classes of outputs can be drawn. Below the line corresponding to the network of figure 3-6 has been drawn.

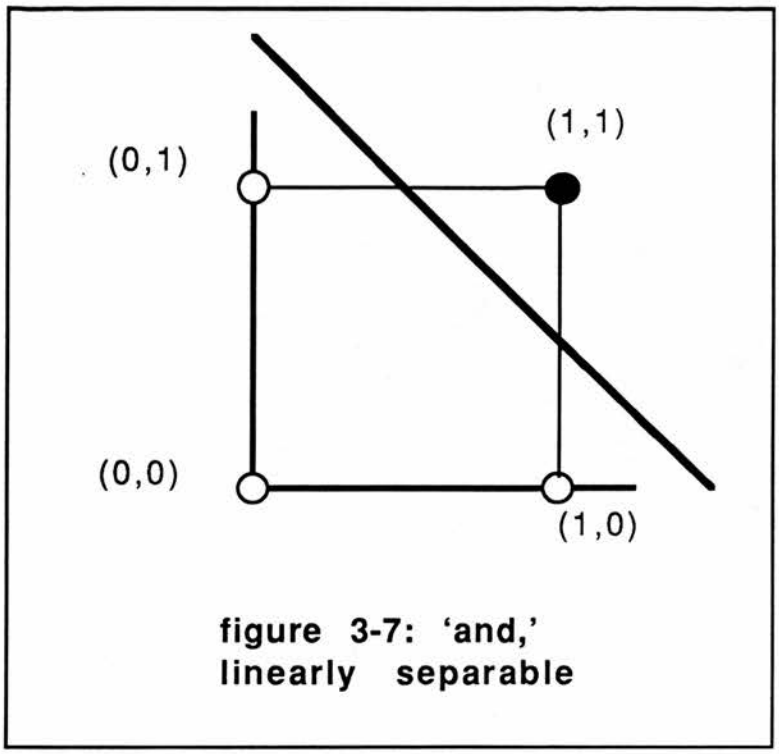


figure 3-7: 'and,' linearly separable

Other functions, such as 'exclusive–or,' are not linearly separable. For a system with two input units and one output unit to be able to compute exclusive-or, the response to stimuli (0, 0) and (1, 1) should be the same (namely 0). But if both input (1, 0) and input (0, 1) have to exceed the threshold value, then it is impossible that input (1, 1) will not exceed the threshold value. Thus the exclusive-or function is not linearly separable, and it cannot be computed by a system like the one in figure 3-6. This problem can be solved by introducing an intermediate (or 'hidden') unit. Figure 3-8 shows the simplest multilayer network which computes exclusive-or for two inputs. The hidden unit is activated only when both input units are activated at the same time (input 1, 1). In this case, the hidden unit sends strong (-2) inhibition to the output unit, and therefore the output unit's threshold value (0.5) is not exceeded. The system of figure 3-8 produces output 1 to inputs (1, 0) and (0, 1), and output 0 to inputs (0, 0) and (1, 1). So its parameters (weights and thresholds) are able to compute exclusive-or.

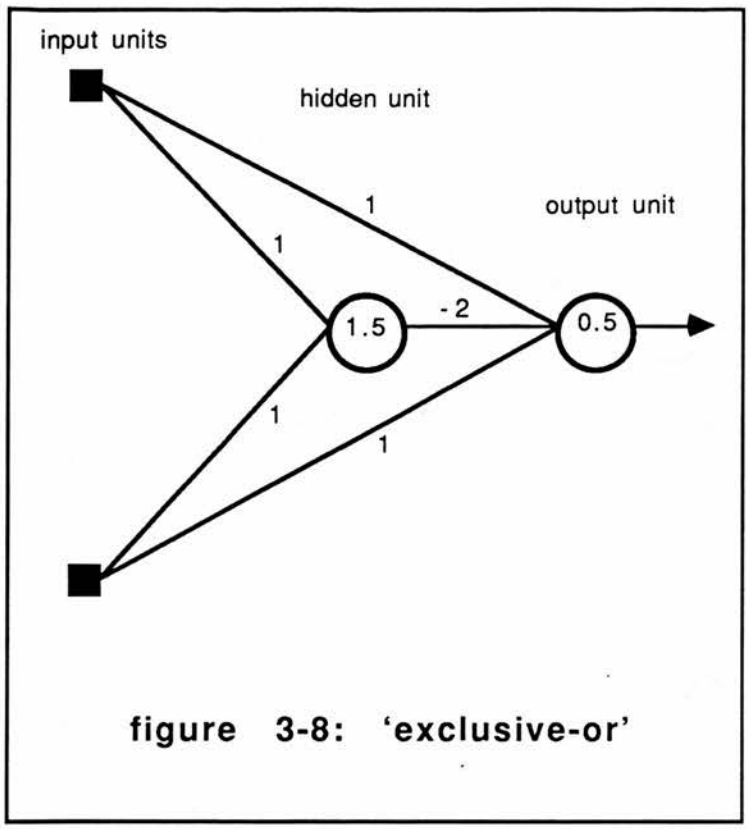


figure 3-8: 'exclusive-or'

It was said in earlier sections that early neural network researchers knew that a solution to some of the problems of single-layer networks was possible by introducing a layer of hidden units between the input and the output layers (e.g. Hawkins, 1961, pp. 45-47). The trouble with that, however, was that those researchers could not find a connection modification procedure equivalent to the ones for single-layer systems (i.e. with interesting convergence properties) for multilayer networks.

Minsky and Papert (1969, ch. 3) showed that the order (see definition of 'order' above) required for a single-layer perceptron like the one in figure 3-4 to compute parity was the whole retina, that is at least one association unit has to receive connections from all the input units. But if one association unit has to 'look at' all the input units in the retina then the computation realised by the perceptron is not based on a combination of *local* information. 'Conjunctive localness,'

defined by Minsky and Papert as the criterion for efficient neural network computation, is lost here.

The second problem studied by Minsky and Papert (1969) which I will discuss here is the 'connectedness' (or figure-ground) problem . It was shown in section 3.2 that this issue worried Rosenblatt significantly. The connectedness predicate consists of saying whether a set of activated retina points belong to the same object (i.e. are connected to each other) or not. The input pattern appearing in the retina of figure 3-9 is connected (all the activated input units belong to the same object).

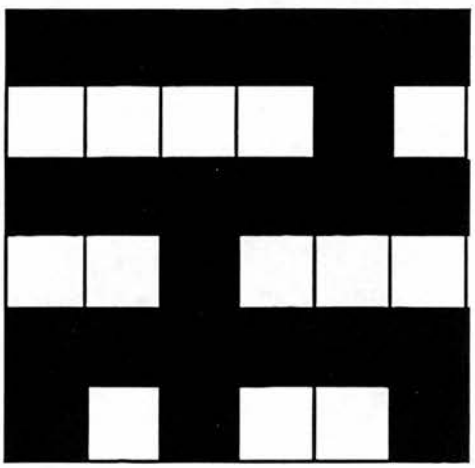


figure 3-9: connected

Minsky and Papert (1969, ch. 5) claimed that the order required for the perceptron of figure 3-4 to compute the connectedness predicate exceeded practical and acceptable limits too. In other words, this order grew 'arbitrarily large' as the input retina grew in size. (It could not be worse than parity, because parity was the worst case, with at least one association unit having to receive connections from all the points of the retina).

> "An instructive example is provided by $\psi_{connected}$ [the connectedness predicate] . . . *Any* perceptron for this predicate on a 100 x 100 toroidal retina *needs* partial functions that *each* look at many hundreds of points! In this

case the concept of 'local' function is almost irrelevant: the partial functions are themselves global." (Minsky & Papert, 1969, p. 17)

"Of course, if some φ [association unit] is allowed to look at *all* the points of R [retina] then $\psi_{connected}$ can be computed, but this would go against any concept of the φ's as 'local' functions." (ibid., p. 8)

Minsky and Papert showed that the order required to compute parity and connectedness with a perceptron was not finite, i.e. that it increased with the size of the perceptron's retina. Igor Aleksander and Helen Morton (1990, p. 41) recently made an interesting comparison in this respect between the perceptron and a conventional computer program.

"Minsky and Papert's [1969] central argument is that perceptrons are only good if their order remains constant for a particular problem irrespective of the size of the input 'retina'. This is similar to the requirement that a program in conventional computing, such as a routine for sorting a list of numbers, should be largely invariant to the size of the task. It is accepted that such a program might need to be given the length of the list as input data, but it would be of little use if it had to be rewritten for lists of different lengths."

Minsky and Papert (1969) also studied the connectedness problem in perceptrons with a different kind of restriction in the input-to-association connections. Instead of order-limited, these were diameter-limited perceptrons. Each association unit was only allowed to 'look at' a circle-shaped limited area of the retina (see figure 3-11). Minsky and Papert showed that diameter-limited perceptrons could not recognise 'efficiently' the connectedness of a figure. The simplified version of their proof (ibid., pp. 12-14) is easy to visualise. They studied the possible responses of the diameter-limited perceptron to the four inputs appearing in figure 3-10.

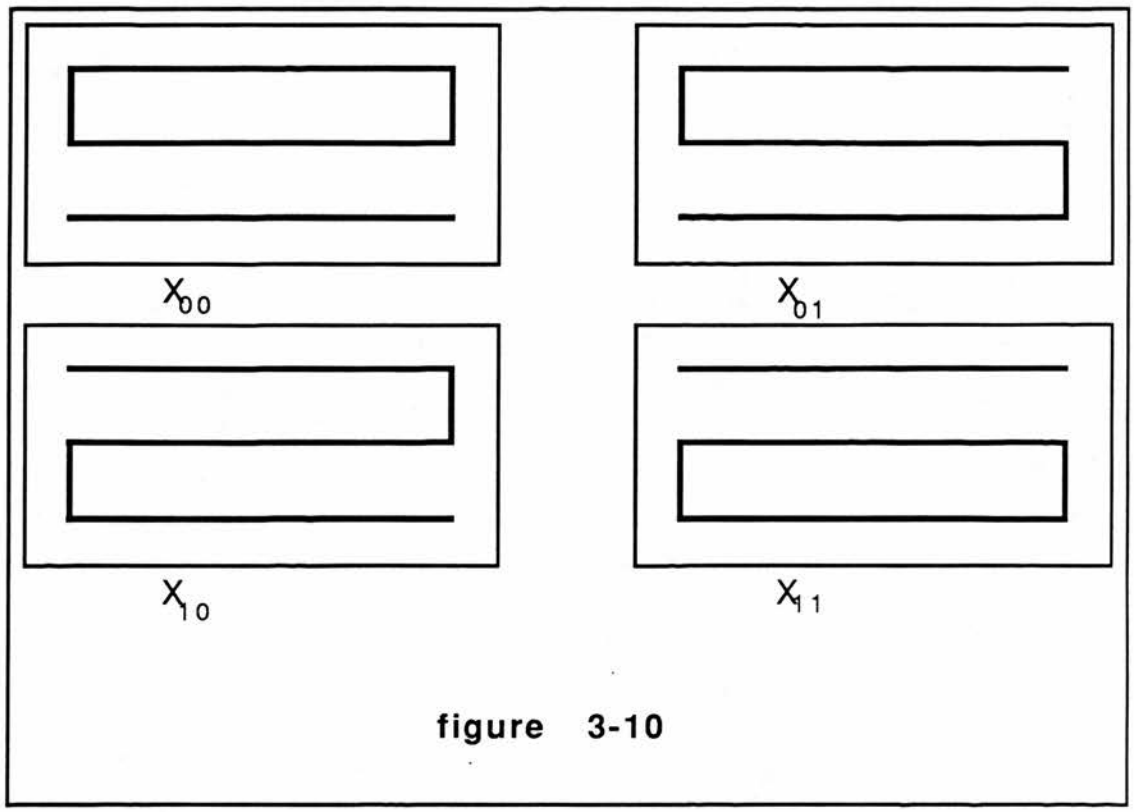


figure 3-10

Minsky and Papert assumed that the diameter-limited input units of the perceptron were structured in groups, as figure 3-11 shows. In this figure, the diameter-limited perceptron has been presented input x_{10} (see figure 3-10).

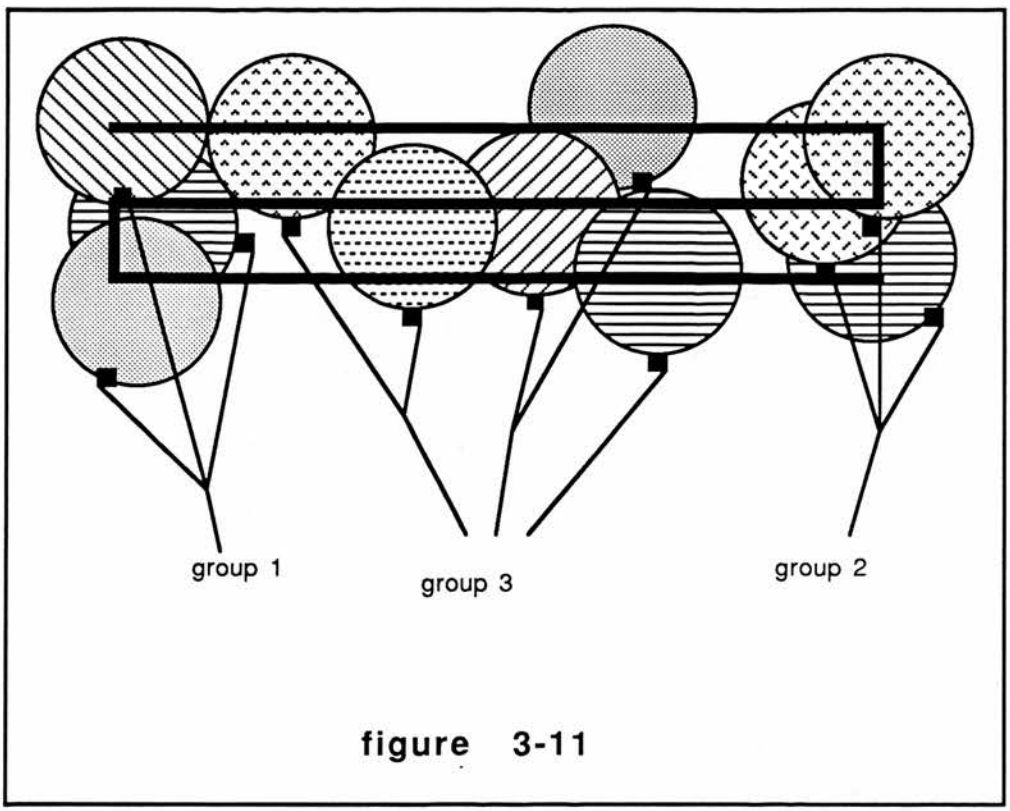


figure 3-11

Assuming that the rest of the perceptron appearing in figure 3-11 is like the one in figure 3-4, the the output unit would fire (the output would be 1) if:

$$\left[\sum_{\text{group 1}} v\ w + \sum_{\text{group 2}} v\ w + \sum_{\text{group 3}} v\ w - \theta \right] > 0$$

Otherwise the output would be 0. Thus this system can classify objects into two clusters (it has one binary output unit). Minsky and Papert showed that this system could not classify the inputs appearing in figure 3-10 as connected or unconnected. For figure X_{00} the sum of the activations of the three groups must be negative (the figure is not connected). Afterwards, if the machine is going to classify X_{10} correctly, the value of group 1 has to be increased to make the total activation greater than the threshold (figure X_{10} is connected). On the other hand if, after classifying X_{00}, the machine had had to classify X_{01}, then the value of group 2 would have increased (X_{01} is connected). But in

changing from X_{00} to X_{11} the value of both group 1 and group 2 would have to increase, since both groups observe the same local changes as in the two previous cases. This would make an even greater total activation than in previous cases. Hence if the perceptron is to make the right decision in changing from X_{00} to X_{10} and from X_{00} to X_{01}, then it would classify X_{11} as connected, since both group 1 and group 2 have access to the same local information. But X_{11} is not connected, and therefore the classification would be wrong. Minsky and Papert concluded that connectedness could not be efficiently computed by a perceptron (i.e. by combining local information in parallel).

One important move in Minsky and Papert's (1969, ch. 9) rhetoric was to claim that problems such as parity or connectedness could be easily solved (i.e. could be easily computed) using conventional algorithms in serial computers.

> "The predicate $\psi_{connected}$ seemed so important in this study that we felt it appropriate to try to relate the perceptron's performance to that of some other, fundamentally different, computation schemes . . . We were surprised to find that, for serial computers, only a very small amount of memory was required." (Minsky & Papert, 1969, p. 72)

> "*Many of the theorems show that perceptrons cannot recognize certain kinds of patterns. Does this mean that it will be hard to build machines to recognize those patterns?* No. All the patterns we have discussed can be handled by quite simple algorithms for general-purpose computers." (ibid., p. 227)

Aleksander and Morton (1990, p. 39-40) described some simple algorithms for computing the parity and connectedness of the retinas appearing in figures 3-5 and 3-9 above:

> "(i) Scan the picture points line by line, left to right, starting at the top left–hand corner of the image until the first black square is reached. (The blobs are assumed to be black on a white background.) (ii) Mark this square and find all its black nearest neighbours. Then mark these neighbours and all their nearest black neighbours and so on

until no new black elements can be found. (This marks all the elements of a blob.) (iii) Remove all the marked elements (by turning them from black to white: this removes the blob.) (iv) Scan the image again and if any black element is found, the image is not connected. The parity task is executed just as easily: the scan-and-remove procedure can be used as before, it then becomes merely a question of counting the number of times the blobs have to be cleared. If this number is even, the image possesses parity."

By showing that parity and connectedness could be quite easily computed by conventional algorithms in serial, von Neumann computers, Minsky and Papert were 'mobilising' two very powerful allies in their favour: the symbol-processing approach to AI and the digital computer. The introduction of these two allies was very important in tipping the 'balance of power' of the controversy in their favour. It was a strong rhetorical move.

But 'strong' does not mean 'necessarily decisive.' The limitations of Laudan's definition of 'anomalous problem' (see quotation above) can be shown here. A 'Laudan style' account of parity and connectedness as anomalous problems would go as follows. Parity and connectedness were anomalous puzzles for neural computing because they could be solved by another research programme in the same domain (namely symbolic AI with its conventional serial algorithms). But this accounts misses important points. There was nothing 'intrinsically' or 'necessarily' anomalous for neural network research in the problems of parity and connectedness. Neural network researchers realised that this was so, and made their own counter-attacking rhetorical moves.

The interpretative flexibility of the problems of parity and connectedness can be shown with an example. Consider figures 3-12 and 3-13. It is not immediately obvious whether this type of figures are connected or not. Figure 3-12 is not connected; figure 3-13 is connected. Consider now the white background as a figure, and look at the centre of the drawing (this is perhaps a

better example). The object appearing in figure 3-12 is now connected; the one in figure 3-13 is now unconnected. But this is not obvious by looking at the centre of the objects. A 'conscious,' 'serial' process is necessary to determine the connectedness of these figures.

figure 3-12

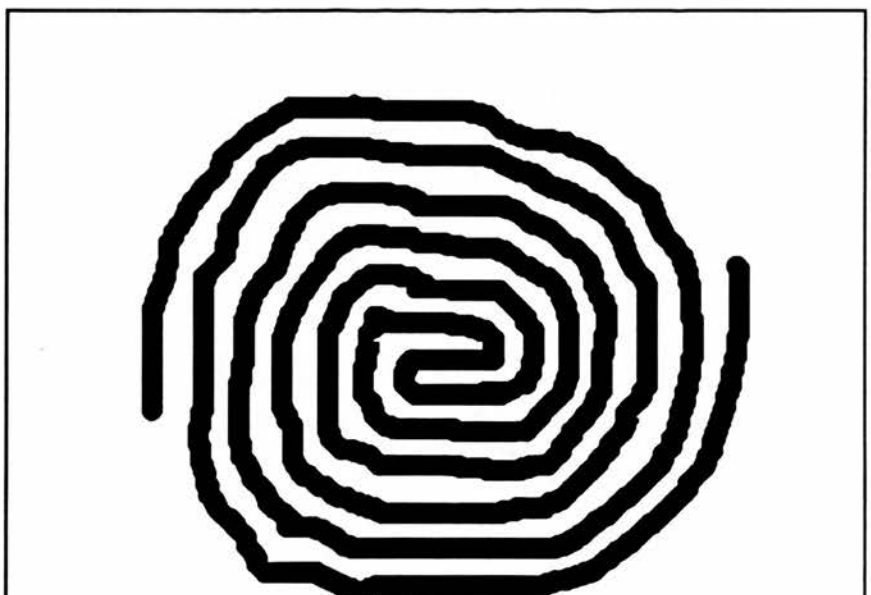


figure 3-13

Early neural network researchers conceded that perceptrons were not very good at recognising parity or connectedness, but — they added — neither are human beings. For an example of the difficulty of the recognition of parity, see figure 3-14: are the number of little black squares odd or even? It is impossible to say without carefully, slowly, and 'serially' counting them.

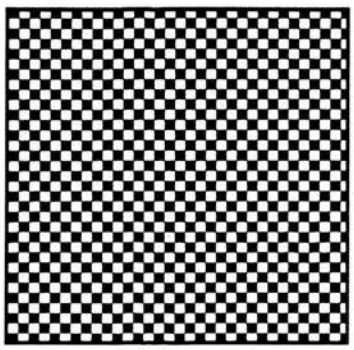


figure 3-14

The 'anomalous' character of the problems of parity and connectedness was open to interpretative flexibility and social negotiation. Perceptron researchers took advantage of this according to their own interests, and replied to Minsky and Papert by indicating that, if one is trying to explain and model human cognitive capabilities, then problems like parity and connectedness are not so anomalous after all. Human beings are not good at recognising parity or connectedness. David Block (1970, p. 517), a mathematician from Cornell University who was a colleague of Rosenblatt in the perceptron project, replied in this way to Minsky and Papert's insistence on the problems of parity and connectedness:

> "Another indication of this difference of perspective [between Rosenblatt and Minsky-Papert] is Minsky and Papert's concern with such predicates as *parity* and *connectedness*. Human beings cannot perceive the parity of large sets (is the number of dots in a newspaper photograph

even or *odd?)*, nor connectedness (on the cover of Minsky and Papert's [1969] book there are two patterns; one is connected, one is not. It is virtually impossible to determine by visual examination which is which. Rosenblatt would be content to approach human capabilities, and in fact would tend to regard unfavorably a machine which went beyond them, since it is human perception he is trying to approximate."

The brain/machine theme has been a constant rhetorical resource throughout the history of neural network research. If the machine (neural network system) works well, then one does not care about the neurobiological or psychological plausibility of the system's architecture. But if one runs into trouble, as perceptron researchers did in this occasion, then one is happy with a machine that is as 'stupid' as human beings. Early neural network researchers were using the tactic of 'triangulation' (Star, 1989a) here. Evidence from different 'local' sources (neural network machines, the brain, human beings' cognitive capabilities) was being used to eliminate uncertainties at the 'global' level (the perceptron or neural network position). Emphasis could be changed from one local area to the other (e.g. from machines to human capabilities) as circumstances required (e.g. when they were having trouble with their machines), so that the global enterprise (neural network research) remained as far from criticism as possible.

Perceptron researchers replied to Minsky and Papert's rhetoric with a different type of rhetoric: triangulation and 'bear with us' (they also used some irony, e.g. 'is the number of dots in a newspaper photograph *even* or *odd?* ' from Block's quotation above). The point was, of course (section 3.4), whether this rhetoric was *strong enough* to contest Minsky and Papert's move. Minsky and Papert selected arguments (parity and connectedness) which were favourable to their position, and Block replied with arguments supporting the neural network position. The 'anomalous' character of a problem is a matter open to interpretative flexibility (the notion of interpretative flexibility was developed by Collins, [1981a, 1983]). Different groups of

scientists, with diverse goals or interests, may interpret an allegedly anomalous puzzle in different ways. Indeed, a group may even see it as 'non-anomalous' at all. Indeed, for Rosenblatt and Block (see quotation above), parity and connectedness were far from being anomalous problems. Quite the contrary: they were 'successful' examples favouring the perceptron approach. Remember Barnes' quotation above: 'What one scientist sees as an anomaly another sees as a puzzle for the same paradigm — even a successfully solved puzzle.'

In previous sections it was shown that early neural network researchers were aware of the difficulties which single-layer perceptrons had in computing predicates like exclusive-or and connectedness well before Minsky and Papert's (1969) study came out. But for them the existence of these problems was not a strong argument against the neural network approach. Single-layer perceptrons were only the first stage in their programme of research. It was shown earlier (see section 3.2) that Rosenblatt had openly admitted the limitations of single-layer perceptrons. In particular, Rosenblatt's insistence on the connectedness problem was almost repetitive. But Rosenblatt and the other early neural network researchers had an approach to the limitations of the perceptron which was remarkably different from that of Minsky and Papert. For Minsky, Papert, and many other symbolic AI researchers, problems like connectedness and parity were decisive arguments against the whole neural network position, and not just against single-layer perceptrons. But for researchers like Rosenblatt, Block, Widrow and others the limitations of single-layer perceptrons were a reason for doing research on more complex perceptrons: systems with more than one layer of modifiable connections, with connections among the units of the same layer, with backward connections, etc.. Neural network researchers emphasised the 'positive' properties of perceptrons (e.g. learning; brain-like, distributed memory; resistance to damage; parallelism). The opposing approaches to (and evaluations of) the limitations of

the single-layer perceptron were made clear by Block (1970, pp. 513-514):

> ". . . The *simple perceptron* (which consists of a set of inputs, one layer of neurons, and a single output, with no feedback or cross coupling) is not at all what a *Perceptron* enthusiast would consider a *typical Perceptron*. he would be more interested in Perceptrons with several layers, feedback and cross coupling . . . The simple Perceptron was studied first, and for it the '*Perceptron* convergence theorem' was proved. This was encouraging, not because the simple Perceptron is itself a reasonable brain model (which it certainly is not; no existing *Perceptron* can even begin to compete with a mouse!), but because it showed that adaptive neural nets, in their simplest forms, could, in principle, improve. This suggested that more complicated networks might exhibit some interesting behavior. Minsky and Papert view the rôle of the *simple Perceptron* differently . . . Thus, what the *Perceptronists* took to be a temporary handhold, Minsky and Papert interpret as the final structure."

'Bear with us' rhetoric (Star 1989a, pp. 137-138) is being used here by Block to respond to Minsky and Papert. For Block, perceptron research simply needed time. Although results could not be delivered yet, neural networks was a 'promising' approach, and its assumptions had to be accepted in the meantime. But time was running out for perceptrons. Many AI-oriented neural network researchers had 'deserted' (so to speak) the neural network position, and developing a powerful enough response to Minsky and Papert's rhetoric was difficult.

Bernard Widrow's account of Minsky and Papert's (1969) 'Perceptrons' is one more example of early neural network researchers' views of Minsky and Papert's book.

> "When I first saw the book, years and years ago, I came to the conclusion that they had defined the idea of a perceptron sufficiently narrowly so that they could prove that it couldn't do anything. I thought that the book was relevant, in the sense that it was good mathematics. It was good that somebody did that, but we had already gone so far

> beyond that. Not beyond the specific mathematics that they had done; but the structures of the networks, and the kinds of models that we were working on were so much more complicated and sophisticated than what they had discussed in the book. All the difficulties, all the things that they could prove that the perceptron couldn't do were pretty much of noninterest, because we were working with things so much more sophisticated than the models that they were studying. The things they could prove you couldn't do were pretty much irrelevant." (Widrow, interview)[50]

A remark by Widrow in this quotation indicates the way to follow in this study of the perceptron controversy: 'the kinds of models that we were working on were so much more complicated and sophisticated than what Minsky and Papert had discussed in the book.'

Neural network researchers of the 1960s were aware of the limitations of the single-layer perceptron, and were studying more complex systems. So the following question would seem decisive: which conclusions about more complex perceptrons did Minsky and Papert (1969) draw from their 're-enacting' study? It is interesting to note that Minsky and Papert made only a few comments about this question. Nevertheless, it will be shown in this and later sections that those comments have been very important throughout the history of neural network research.

On the issue of perceptrons more complex than the single-layer one, Minsky and Papert stated a pessimistic 'intuitive judgement' about the possibility of developing an efficient technique for

---

[50] Widrow and Lehr (1990, p. 1422) claimed recently that the fact that a typical processing unit in a neural network system can only realise linearly separable functions (i.e. has got limited capacity) is an advantage, rather than a limitation. It facilitates generalisation, a very important property of neural networks: "A [neural] network's capacity is of little interest unless it is accompanied by useful generalizations to patterns not presented during training. In fact, if generalization is not needed, we can simply store the associations in a look-up table, and will have little need for a neural network. The relationship between generalization and pattern capacity represents a fundamental trade-off in neural network applications: the Adaline's inability to realize all functions is in a sense a strength rather than the fatal flaw envisioned by some critics of neural networks (Minsky & Papert, 1969), because it helps limit the capacity of the device and thereby improves its ability to generalize."

training multilayer systems. It was said earlier that multilayer systems can realise classifications which single-layer networks are not capable of ('exclusive-or' is the simplest example). In the 1960s it was known that multilayer systems could realise powerful classifications, but adequate algorithms for adjusting the connections of these systems had not been developed. In the rest of this section I look at Minsky and Papert's view of the problem of learning in multilayer networks, and how it affected the development of the perceptron controversy.

Thomas Hughes' (1983) concepts of 'reverse salient' and 'critical problem' can help understand some aspects of the evolution of the problem of training multilayer neural networks. The term 'reverse salient' has its origins in the field of military historiography. In that context, a reverse salient is a section of an advancing military front (represented as a continuous line) which has fallen behind for some (and varied) reasons (ibid., p. 79). For Hughes this metaphor is useful because it refers to a complex situation in which many different factors may intervene, that is a situation shaped by a complex diversity of circumstances and determinants.

Hughes uses the concept of reverse salient in his studies of the evolution of technological systems. Reverse salients are problems which obstruct the development of technological systems. According to Hughes, those problems are obvious to the agents involved in a technological system. The difficulty does not lie in localising a reverse salient, but in giving a satisfactory solution to it. When a reverse salient is defined as a problem that can be solved, then, in Hughes' terminology, it becomes a 'critical problem'. For Hughes, defining reverse salients as critical problems is the key to technological innovation and change.[51]

---

[51] "Reverse salients are obvious weak points . . . in a technology which are in need of further development. A reverse salient is obvious, and creative imagination is not needed to define it. In contrast . . . the definition of critical problems by an inventor does require creative imagination. Critical problems result from the inventor's defining the reverse salient as a problem, or set of problems, that, when solved, will correct the reverse

Hughes applied his reverse salient/critical problem to the study of technological systems, but I see no reason why it cannot be applied to the development of neural network research. The distinction between science and technology is particularly meaningless in the case of neural network research.[52] Or better: it is useful *for rhetorical purposes*. In different moments of the evolution of neural networks 'technological aspects' (i.e. the fact that the machines do something useful, whether brain-like or not) or 'scientific aspects' (i.e. the brain-like character of neural network computation, as compared to symbolic AI) have been emphasised alternatively as rhetorical resources for the legitimisation of neural computing (see comments on triangulation earlier in this section). This tension between 'information technology' and 'brain' goes back to the brain/machine theme in cybernetics, and it is characteristic of all approaches to AI and cognitive science research. Both symbolic AI and neural networks are methodologies for building intelligent machines, and they are also frameworks for the study and modelling of cognitive processes.

But let me come back to the problem of learning in multilayer neural networks. It was seen in earlier sections that this problem had been on the agenda of neural networks researchers long before Minsky and Papert's (1969) study was published. Of course it was not the only problem, but it did receive a very considerable amount of attention. Minsky and Papert's (1969) short comments about multilayer systems emphasised the 'reverse salient' character of this problem. This was an effect of the decisive role played by Minsky and Papert's (1969) criticism

---

salient . . . Defining reverse salients as critical problems is the essence of the creative process. An inventor or applier of science transforms an amorphous challenge — the backwardness of a system — into a set of problems that are believed to be solvable . . . The articulation of a problem often implies its solution." (Hughes, 1983, pp. 22 and 14-15)

[52] Researchers in the sociology of technology have indicated that it is very problematic to separate science and technology (see e.g. Bijker, Hughes, & Pinch, 1987, p. 10). Bruno Latour (1987) insisted in the uselessness of the distinction, and suggested the term 'technoscience.'

in the closure of the perceptron controversy. It can be said that Minsky and Papert helped 'construct' the 'reverse salient' character of the problem of learning in multilayer networks. By focusing only on that aspect of multilayer systems, they emphasised its importance, and so in a sense they 'constructed' a reverse salient for neural network researchers. That reverse salient was far from being developed as a 'critical problem' in Hughes's sense, but it is important to note that its character of reverse salient became even clearer after Minsky and Papert's (1969) study. Indeed it became a challenge for neural network researchers thereafter.

It is worth analysing carefully Minsky and Papert's short comment about the question of learning in multilayer neural networks. Minsky and Papert said little else apart from this pessimistic 'intuitive judgement' on the issue of perceptrons more complex than the single-layer one:

> "The perceptron has shown itself worthy of study despite (and even because of!) its severe limitations. It has many features to attract attention: its linearity; its intriguing learning theorem; its clear paradigmatic simplicity as a kind of parallel computation. There is no reason to suppose that any of these virtues carry over to the many-layered version. Nevertheless, we consider it to be an important research problem to elucidate (or reject) our intuitive judgement that the extension is sterile. Perhaps some powerful convergence theorem will be discovered, or some profound reason for the failure to produce an interesting 'learning theorem' for the multilayered machine will be found." (Minsky, 1969, pp. 231-232)

This statement, which would have great 'symbolic' importance in the re-emergence of neural computing in the 1980s, is a 'monument' to rhetoric. Minsky and Papert start with diplomacy: 'the perceptron has shown itself worthy of study;' 'it has many features to attract attention.' Minsky and Papert 'assimilate' some aspects of the position they criticise. These 'positive' statements are also a consequence of the considerable research effort that they had dedicated to 're-enacting' the perceptron, so

that the controversy could be settled once and for all. It was a costly process in terms of time and research effort, and therefore Minsky and Papert had to stress the positive aspects of their work. But they quickly came to the attack (after all, the purpose of the study was to 'kill' the perceptron): 'there is no reason to suppose that any of these virtues carry over to the many-layered version.' In other words, there is no reason to 'bear with them,' 'them' being neural network researchers insisting that results would come from (that solutions would be delivered by) the study of multilayer systems.

Minsky and Papert staged their challenge carefully. They 'took out an insurance policy' (using Latour's [1987, p. 55] words) against possible developments in the future, as though they thought that a learning algorithm could be developed in the future (who knows, stubborn neural network enthusiasts may keep trying!). Their insurance policy against possible future developments was: 'we consider it to be an important research problem to elucidate (or reject) our intuitive judgement that the extension is sterile;' 'perhaps some powerful convergence theorem will be discovered.' These are very carefully staged statements. One could not 'blame' them years later for having said that some kind of 'efficient' learning in multilayer systems was totally impossible or impractical.

But the interpretative flexibility of this comments is, by all standards, much greater than that of the problems of parity and connectedness. Who should be believed? Neural network researchers 'promising' results, or Minsky and Papert's pessimistic 'intuitive judgement' that work on multilayer systems 'will be sterile?' An 'intuitive judgement' does not seem very powerful evidence in favour of a position.

It was pointed out in chapter one that, in principle, interpretative flexibility and controversy can go on for ever. But of course 'practice' is different from 'principle:' interpretative flexibility is reduced and controversies are closed. As Latour (1987) pointed out, 'practice' is very much a matter of power,

and it is not easy to be able to gather enough resources and to enrol enough allies so that the rhetoric (in Latour's sense: from weaker to stronger) of the opposition can be matched. In this section I have not only shown that the neural network position could have been defended in principle, but also that neural network researchers did actually defend their position with their own rhetoric. There was nothing 'necessarily compelling,' 'intrinsically superior,' or more 'rational' in Minsky and Papert's criticism of neural networks than in neural network researchers' defense of their approach.

In the next section I examine how, in spite of Rosenblatt and others' efforts, the plausibility of the neural network position (its power to contest Minsky and Papert's challenge) was increasingly reduced until the perceptron controversy was closed.

In this section I have studied Minsky and Papert's (1969) strong move in the perceptron controversy. They re-enacted the perceptron with a view to showing, clearly and decisively, its flaws and limitations. Minsky and Papert's move was one from weaker to stronger rhetoric. And they mobilised powerful allies in their favour, such as symbol-processing AI and the von Neumann computer (remember how easily parity and connectedness can be computed serially). It is usually thought that Minsky and Papert *showed* that the perceptron had so many problems that it was not worth doing further research in neural networks. But I have shown here that this conclusion was not clear at all before the perceptron controversy was closed (section 3.4). That view is the result of the closure of the controversy, not its cause. Before the closure, Minsky and Papert's results were very much open to interpretative flexibility. And not only in principle, but also in practice. In fact, neural network researchers tried to use that interpretative flexibility in their favour in order to launch a counter-attack. In the next section I discuss how the controversy was finally closed.

3.4 Closure of the controversy

In this section I discuss the the effect of Minsky and Papert's (1969) critical study on the closure of the perceptron controversy. I study the process through which the interpretative flexibility of Minsky and Papert's arguments was closed against the neural network position. I analyse how the response of the Rosenblatt camp to Minsky and Papert's criticism was not enough to contest the interpretation of that criticism as showing that neural networks (as a whole) were not worth pursuing. Rosenblatt and his colleagues could not mobilise a powerful enough rhetoric to contest Minsky and Papert's challenge. They failed to enlist key allies such as the funding agencies, and they were increasingly isolated. I also show that the linkage between the position against neural networks and powerful actants such as symbolic AI and digital computer technology was very important for the closure of the controversy.

In principle, scientific controversies can always go on (see chapter one). The point is then to analyse how the plausibility of the positions involved in a controversy evolves, and how the in-principle possibility of going on arguing is in practice reduced and controversies are closed. In this section I analyse the process of closure of the perceptron controversy. This process went from the mid-1960s to the publication of Minsky and Papert's (1969) 'Perceptrons' study. It was seen earlier that the crisis of neural computing had reached 'quite worrying' (for neural network researchers) proportions by 1965-1966. Minsky and Papert's (1969) study was the 'last push,' so to speak, for the closure of the controversy. Nonetheless, it is important to note that many of the arguments elaborated in detail in Minsky and Papert's (1969) important book were well known by the mid-

1960s, and that they had affected the crisis of neural network research long before the book came out.

In the previous section I used Latour's concept of 're-enacting' to study the development of Minsky and Papert's (1969) study. It was shown that Minsky and Papert decided to make a decisive move in the controversy, and embarked on their 'Perceptrons' project. It was a costly process in terms of time and research effort, but Minsky and Papert were successful in getting a 'last word' effect in the controversy. It will be shown in this section that the last word was theirs. Nobody was powerful enough to contest their claims. Conclusion: they had 'won' the battle, the perceptron controversy was over (at least for the time being). A 'back box' had been created.

Some questions can be asked at this point: Why was it the last word? Why were neural network researchers not powerful enough to contest Minsky and Papert's challenge? But let me start from another question: who could have responded to the challenge? In sections 2.3 and 3.2 I studied the development of the main early neural network projects. After what was said in those sections, part of the answer to the above question if simple: many neural network researchers had simply abandoned their 'battle positions' by the time Minsky and Papert's (1969) heavy attack was published. So they were hardly in a position to counter-attack.

By the time Minsky and Papert's (1969) study came out, Widrow had already been working on applications of the LMS learning algorithm and the adaline (first developed within neural networks) to adaptive signal processing engineering problems (like antennas and telephone systems) for a few years. These applications were being rather successful and, after the troubles Widrow had had with neural network machines in the mid-1960s, he did not consider the possibility of trying to respond. His reaction to Minsky and Papert's (1969) book was critical (remember his words [Widrow, interview] in the previous section: 'All the difficulties, all the things that they could prove that the

perceptron couldn't do were pretty much of non-interest, because we were working with things so much more sophisticated than the models that they were studying. The things they could prove you couldn't do were pretty much irrelevant'). But coming back to the controversial field of neural networks was not considered practical by Widrow and his colleagues.

A second group which could (in principle at least) have responded to Minsky and Papert's (1969) move was the SRI group. But their neural network 'battle positions' were also empty by the time 'Perceptrons' was published. The SRI researchers abandoned neural network research when funding for Minos run out in 1966. At that time continuation of neural network research was not considered practical by most of them. Furthermore, they were no longer in the 'opposition.' They were now in a sense 'allies' of Minsky. They were carrying out leading research in robotics within the wider umbrella of symbolic AI. (Needless to say, the use of military terminology here is not evaluative.)

The third group which could in principle have responded to Minsky and Papert's (1969) challenge was of course Frank Rosenblatt's. Rosenblatt was the 'symbolic leader' of the neural network position in the perceptron controversy, and therefore his reaction to Minsky and Papert's criticism was especially important. He was the most interested person in replying to Minsky and Papert's (1969) attack. After all, that attack was the consequence of Minsky and Papert's re-enacting of his own machine: the perceptron. Minsky and Papert's objective in their 're-enacting' project was very much the criticism of Rosenblatt's perceptron (remember rhetoric like 'kill the perceptron'). So the question is obvious: Why did Rosenblatt not contest Minsky and Papert's 'heavy attack move'? The answer can only be one: because he was not powerful enough, because he was not successful in enlisting enough allies and actants, and because he could not gather or mobilise enough resources to launch a powerful counter-attack.

One important aspect of the perceptron controversy was the 'funding issue.' The question of funding was among Minsky and Papert's motivations for starting their 'Perceptrons' project (see section 3.1). Remember some remarks by Minsky and Papert from section 3-1: 'They were trying to get money to build bigger machines;' 'part of our drive came from the fact that funding and research energy were being dissipated on misleading attempts to use connectionist methods in practical applications.' Remember also, from section 3.1 on the 'heat of the controversy,' that much of the popularity of the perceptron goes back to the funding of Rosenblatt's project by the Office of Naval Research (ONR) and the 1958 press release. In the 1960s, as the perceptron controversy increased, Rosenblatt failed to enlist key allies such as ONR and the Advanced Research Projects Agency (ARPA, now DARPA).

Frank Rosenblatt died in a boat accident in 1971, and so did later his colleague David Block.[53] Information about funding for the perceptron project throughout the 1960s (after Mark 1) is not easy to obtain. I was able to talk with two people who were at ONR at the time: Marshall Yovits and Marvin Denicoff. Yovits was at ONR in the late 1950s and early 1960s, and was responsible for the funding for Rosenblatt and his group. Denicoff was at ONR from the mid-1950s to the early 1980s. This is important, because ONR was the main (and probably the only) source of financial support for Rosenblatt over the years.

Rosenblatt's relationships with the US military were never easy. Rosenblatt had some kind of involvement in the peace movement at the time of the Korean War, and as a consequence he never got a security clearance from the military. Marshall Yovits (interview) told me about this:

---

[53] Rosenblatt died in a boat accident while sailing with two students from Cornell University in the Chesapeake Bay near Easton, MD, on his 43rd birthday (see New York Times, July 13). David Block died a few years ago of a heart attack (Nilsson, interview).

> "Rosenblatt was somehow involved in some peace movements. This was during the Korean War, and as a consequence he was never able to receive a security clearance. In those days, after World War II — it was the era of McCarthy — there was much emotional anticommunist concern. If you were accused of being a 'left-winger' you'd lose your security clearance, and so on. This let up somewhat after McCarthy died, but not a lot, and it went on through the Korean War. Remember that this was during the Eisenhower presidency when even J. Robert Oppenheimer lost his security clearance. Rosenblatt was not in favour of the Korean war, and he was involved with some sort of a peace movement. As a consequence, he never did receive any sort of security clearance. In those days things were very tight, but we were able to work with him nevertheless."

Yovits was at ONR in the late 1950s and early 1960s, when interest in the perceptron was at its highest level.[54] Later Rosenblatt was not so lucky. Later in the 1960s he was not able to get funding for a 'large' perceptron research project. Another way of putting this (confirmed by researchers who were working on neural networks at the time) is that Rosenblatt's financial support was cut at some point. It is difficult to establish the exact date at which this happened, but it is clear that Rosenblatt was unable to enlist and convince key allies like ONR and ARPA in the mid- and late 1960s.

> "The Office of Naval Research, which I believe was Rosenblatt's main source of support, and maybe his only source of support, rarely supported big projects. Our programs were small, and supported key scientists to get their work going. Rosenblatt was never able to get the big

---

[54] One indicator of the level of activity in neural network-like research at the time, and ONR's involvement in it, are the three important conferences on 'self-organisation' which were held in 1959, 1960, and 1962. (proceedings published as [Yovits & Cameron, 1960]; [von Foerster & Zopf, 1962]; and [Yovits, Jacobi, & Goldstein, 1962] respectively). Yovits himself was one of the organisers of these conferences, and contributions presented there are a good sample of the work which was being carried out in neural networks and related fields in the late 1950s and in the first years of the 1960s. The variety of approaches to the brain/machine cybernetics issue in those conferences was pointed out in section 2.1.

dollars that were needed in order to build the machinery that he thought had to be built. But even if he had been able to get the dollars, he lived in the wrong period; the means of implementation were not there. [Even] if more money had been available, I'm not quite sure what would have been done with it. A large machine in my opinion would have been pointless. As far as I recall, at that time, investigators just began to lose interest in the neural net field. The attitude was, we had shown that Rosenblatt's device works in a simple way, but it didn't really have any future. This was before VLSI." (Yovits, interview)

In the above quotation, Yovits seems to be implying that Rosenblatt was not successful in convincing someone else, not only ONR ('our programmes were small'). That someone else could well be ARPA. 'Big dollars' for AI-like research in the 1960s (and after) came from ARPA. Marvin Denicoff confirmed this. Denicoff was at ONR after Yovits left, and he was involved in funding projects in AI and related fields, sometimes in partnership with ARPA (so he was also well informed about ARPA's activities).

"At that time [in the 1960s], the Office of Naval Research had funds at the level of, $ 40 or 50 K. ARPA was able to fund hundreds of thousands, or even millions. Rosenblatt never attracted that kind of money, because he wasn't offering a large pay-off. By pay-off I mean not in the scientific sense, but in the application sense, world problem solving. Again, his work was much more, I would say, traditional science. The Office of Naval Research never gave him the kind of money that he really required, and he was not successful in getting the money from the Science Foundation or from ARPA. One can draw the conclusion that if he had had the money he would have made even greater progress. That's too easy an answer, because it doesn't always follow that large amounts of money make the difference . . . Well before the Minsky and Papert [1969] book came, he [Rosenblatt] was not successful in attracting more money, that I know for a fact . . . Again, each thing has its moment in time, that's another point. I will give you one theory that I have. For any funding programme, whatever it is, within a few years you've got an 80 or 90% of the progress that you will ever get. All of the bright ideas come out very quickly. From there on, the hill climbing is

> very steep and very slow . . . The money very seldom grows, it keeps getting redistributed. So as each new exciting field comes along, something else gets sort of pushed aside a little bit, and then there are a wing of people who can claim (I am not saying that all of that is unjustified): 'what a serious mistake they made.' If you knew how many times I've heard 'if I had had one more year, one more year, I would have done it, just one more million dollars'. . ." (Denicoff, interview)

Denicoff pointed out that Rosenblatt was not offering a big pay-off in terms of applications. His words can be related to the general context of science policy in the United States in the 1960s. From the late 1950s to the mid-1960s funding for science in the United States had a period of unprecedented growth (Dickson, 1988, pp. 5-7). The beginning of this growth goes back to the post World War II period, but the peak in the growth rate corresponds to the late 1950s and early 1960s. The opening of the space race with the launching of the Sputnik satellite in October 1957 by the Soviet Union was the catalyser. Support for science increased to unprecedented levels. ARPA itself was created within the Defense Reorganization Act of 1958, a reaction to the Sputnik launch.

ONR's support for Rosenblatt's perceptron in the late 1950s and early 1960s came within this context. At that time ONR earned a reputation of working without worrying too much about the pay-off in terms of applications.

> "[In the early 1980s] Many scientists looked back with nostalgia at the postwar period when, taking their lead from the organization of the wartime Manhattan Project, agencies such as the Office of Naval Research provided generous funding for universities with virtually no strings attached. This approach is compared unfavorably with the many social demands on the research community introduced in the late 1960s and 1970s — in particular, the demand for direct social accountability (illustrated, in the case of military research, by the requirements of the Mansfield Amendment . . .)." (Dickson, 1988, p. 113)

The situation did not last long, however, and the growth rate in funding for scientific research decreased significantly from about 1965. Dickson (1988, pp. 5-6 and 123) indicated this:

> ". . . Political enthusiasm, grounded in the success of the Manhattan Project, spurred by the shock of the Russian Sputnik, and reaching its apogee during the Kennedy administration, provided scientists with both lavish financial support and high social status. This period was followed, from the mid-1960s, by a stage of questioning and doubt, when more direct payoffs were asked . . . The decline had in fact started in 1965, well before the Mansfield Amendment was passed."

The growing concern in the Defense agencies with the applications of the research they funded was reflected in the 'Mansfield Amendment' to the Pentagon's budget for 1970, which stated explicitly that "research should be supported only if it could demonstrate direct relevance for some military need" (ibid., p. 30).

This general context of science policy from the mid-1960s onwards did not favour Rosenblatt's chances of being funded. In the 'Tribute to Frank Rosenblatt,' held in July 1971 after Rosenblatt's death, Richard O'Brien, head of the Division of Biological Sciences of Cornell University (where Rosenblatt was working at the time) made the following comment (Congressional Record, 1971, p. 3) in his speech:

> ". . . It was only a few years ago that he enjoyed hundreds of thousands of dollars a year in research grants, from agencies that thought his work was worth doing, and he was a victim of the Mansfield amendment, and within a few years that money melted like summer snow and soon he had very little left in the last few months."

In the mid- and late 1960s Rosenblatt worked on perceptrons and (at least) two other projects: a memory transfer project and an astronomy project. He carried out several memory transfer experiments with rats (e.g. Rosenblatt et al., 1966), and he was

also interested in a problem in astronomy, namely photometric detection of extra-solar planetary systems (Scattergood, personal communication). But it is important to emphasise that Rosenblatt continued doing perceptron research. Information about his later perceptron projects is deeply buried, but several of my interviewees and researchers who worked with or close to him at that stage confirmed that Rosenblatt was working on a perceptron called 'Tobermory' until his final years. This is how Richard O'Brien described Frank Rosenblatt's latest perceptron project:

> ". . . Frank became interested in a massive expansion of fabricated perceptrons, as follows (I am sure you are aware that, until his final years, he was working simultaneously upon computer simulations of perceptrons and physical assemblage of them. An enormous assemblage filled a large room in the Langmuir Laboratory where he worked). He wanted to create a synesthetic perceptron called Tobermory, named after the infamous cat in the short story by Saki. Tobermory was going to be able to perceive a mouse running across the room and say (out loud): 'I see a white object with a long tail making a squeaking noise and it must be a mouse.' Thus Tobermory would be able to see, hear, and speak, and to synthesize all three elements appropriately." (Richard D. O'Brien, personal communication)

Unlike Widrow or the researchers at SRI, Rosenblatt had not abandoned perceptron research when Minsky and Papert's (1969) study was published. The question therefore is: what was the reaction of Rosenblatt's camp to Minsky and Papert's criticism?[55]

One important factor is that, as I said before, in the mid- and late 1960s Rosenblatt was unsuccessful in enlisting key allies such as the funding agencies. DARPA, in particular, had already made its choice in favour of symbolic AI, and they were not

---

[55] Information on Rosenblatt's *personal* reaction to Minsky and Papert's (1969) book is very difficult to obtain, but I am following several lines of enquire in this respect that could lead to more information.

interested in neural networks. ONR continued to fund Rosenblatt's small-scale projects (like his research on memory), but they did not sponsor the larger-scale perceptron projects that Rosenblatt wanted to carry out for two reasons. One is that apparently funding for large-scale projects was not their 'style;' the other (more important) one is that they did not believe in the usefulness of Rosenblatt's perceptron projects (see quotations by Yovits and Denicoff above).

Rosenblatt was on the losing side of the controversy, and was pretty much isolated in his laboratory room ('full of assemblage') working on his Tobermory perceptron project. He was not in a position to enrol other research groups. The times when he convinced (enroled) Rosen and his colleagues at SRI to work on perceptrons were over. Both the SRI group and Widrow's group had abandoned neural networks. And perceptrons had important problems which had not been solved, like training multilayer systems (see section 3.2). Rosenblatt and colleagues were having great difficulties in gathering and mobilising enough resources to make a counter-attack against Minsky and Papert's (1969) criticism.

David Block's (1970) review of Minsky and Papert's (1969) book remains the 'official' response of the Rosenblatt camp to Minsky and Papert's (1969) 'heavy attack.' Parts of Block's response were discussed in the previous section. It was based on two main points. On the one hand, Block accused Minsky and Papert of 'controlling the focus of the debate' (using Star's [1989, pp. 145ff] term). He criticised them for having focused on the single-layer perceptron which was:"not at all what a Perceptron enthusiast would consider a typical Perceptron" (Block 1970, p. 513). See also: "what Perceptronists took to be a temporary handhold, Minsky and Papert interpret as the final structure" (ibid., p. 514).

Furthermore, Block attacked at the point where Minsky and Papert's argument was weakest (and its interpretative flexibility greatest): more complex perceptrons "with several

layers, feedback, and cross coupling" (ibid., p. 513). In his strongest attacking move, Block tried to mobilise as many allies as he could, and relied heavily on 'bear with us' and 'reference to the unknown' debating tactics. His response to Minsky and Papert's pessimistic 'intuitive judgement' on multilayer perceptrons was an appeal to the promising side of perceptron research:

> "Work on the four-layer *Perceptrons* has been difficult; but the results suggest that such systems may be rich in behavioral possibilities, once the mathematical tools become available for analyzing them (cf. Rosenblatt (1960), (1964), Block, Knight and Rosenblatt (1962), Konheim (1963)). Even more suggestive are the multilayer machines with feedback (the *C*-systems and *F*-systems of Rosenblatt (1967)). The models studied extensively by Grossberg (1967-1969), although differing from the perceptron in several respects (continuous variables, instead of discrete; linear, instead of a step-thresholding function, etc.) are nevertheless much closer to the spirit of Rosenblatt's Perceptron than the work under review [Minsky & Papert, 1969]. The same can be said of other brain models, such as those of Kabrisky (1966) or Baron (1970a), (1970b). From this point of view, the potential capabilities of Perceptrons are still mostly unexplored." (Block, 1970, pp. 516-517)

Block mobilised many references to back his point — a point that he undoubtedly though was of great importance. Latour (1987) studied this kind of rhetorical move.[56] By mobilising in one's favour as many references as one can, the proponent of a claim makes him or herself stronger, because the dissenter has now to contest not one, but many scientific texts (of course the proponent has to be careful: the dissenter may also try to show that some of the referred texts do not 'really' support the

---

[56] See for example: ". . . Attacking a paper heavy with footnotes [references in this case] means that the dissenter has to weaken each of the other papers, or will at least be threaten with having to do so, whereas attacking a naked paper means that the reader and the author are of the same weight: face to face. The difference at this point between technical and non-technical literature is not that one is about fact and the other is about fiction, but that the latter gathers only a few resources at hand, and the former a lot of resources, even from far away in time and space" (Latour, 1987, p. 33).

proponent's view, and may so try a counter-attack). It is interesting to note that, perhaps for the first time, Stephen Grossberg's work is referred to in a paper from Rosenblatt's group. Grossberg started his neural network research in the late 1960s. He was one of the researchers who continued doing neural network research in the 1970s, and he is today one of the leading members of the neural network community.

But, in spite of Block's efforts, the interpretative flexibility of multilayer perceptrons was closed in favour of Minsky and Papert's view. There was nothing 'necessarily superior' in Minsky and Papert's 'intuitive judgement' about multilayer perceptrons (see section 3.3) as compared to Block's conclusion that 'the potential capabilities of Perceptrons are still mostly unexplored' (see quotation above), but Block and Rosenblatt were on the losing side.

The emergence and institutionalisation of symbolic AI (see section 2.5) was a very important factor for the resolution of this interpretative flexibility and uncertainty. That emergence and institutionalisation was a 'demonstration of power' by researchers opposed to the neural network approach. It was seen in section 3.3 that, in showing the limitations of perceptrons, Minsky and Papert enroled powerful allies such as symbolic AI and the digital computer (remember the issues of parity and connectedness). The 'promising' side of neural computing, as underlined by Block in his response to Minsky and Papert, was not enough to contest Minsky and Papert's position. With symbolic AI's institutionalisation well under way, backed by key allies such as ARPA, and connected to developments in digital computer technology, the Rosenblatt camp had little room to manoeuvre. Minsky and Papert's (1969) criticism was being widely interpreted as showing the uselessness of neural networks as a whole (more on this below), and Rosenblatt and colleagues were unable to stop Minsky and Papert's 'last word effect:' the perceptron controversy was being closed against the neural network position.

The closure of the perceptron controversy was beneficial for the development of symbolic AI. The widely held belief that there was no credible alternative to symbol-processing helped legitimise the institutionalisation of the symbolic approach. Minsky and Papert's (1969) criticism of neural networks was the final 'push' to the crystallisation of the consensus in favour of the 'adequacy' of the symbolic approach and the lack of credibility of neural networks as an alternative approach to AI. Newell and Simon used this 'lack of a serious alternative' as a rhetorical tactic in the important paper (1976) on the foundations of symbolic AI:

> "The principal body of evidence for the symbolic hypothesis that we have not considered [so far in this paper] is negative evidence: the absence of specific competing hypotheses as to how intelligent activity might be accomplished — whether by man or by machine." (Newell & Simon, 1976, p. 50)

The origin of this 'lack of a credible alternative' view goes back to the crisis of neural network research and the closure of the perceptron controversy. These events could well be the 'marker event for the end of the process of emergence of symbolic AI' that Allen Newell (1983) is looking for in his historical study of the evolution of AI. Newell claims (quite in accordance with what has been said here) that the process of emergence of symbolic AI was 'essentially complete by 1965,' although he cannot find a marker event. I would like to make him a suggestion: the 'marker event' he is looking for is the crisis of early neural network research.

> "Through the early 1960s, all the researchers concerned with mechanistic approaches to mental functions knew about each other's work and attended the same conferences. It was one big, somewhat chaotic, scientific happening. The four issues I have identified — continuous versus symbolic systems, problem solving versus recognition, psychology versus neurophysiology, and performance versus learning — provided a large space within which the total field sorted itself out. Workers of a wide combination of persuasions on

these issues could be identified. Until the mid-1950s, the central focus had been dominated by cybernetics, which had a position on two of the issues — using continuous systems and orientation towards neurophysiology — but no strong position on the other two . . . The emergence of programs as a medium of exploration activated all four of these issues, which then gradually led to the emergence of a single composite issue defined by a combination of all four dimensions [symbolic, problem solving, psychology, performance]. This process was essentially complete by 1965, although I do not have any *marker event*." (Newell, 1983, p. 201, emphasis added)

Later Newell points out at one more 'issue':

". . . Most pattern recognition and self-organizing systems were highly parallel network structures. Many . . . were modelled after neurophysiological structures. Most symbolic-performance systems were serial programs. Thus, the contrast between serial and parallel (especially highly parallel) systems was explicit during the first decade of AI. The contrast was coordinated with the other four issues I have just discussed." (ibid., p. 202)[57]

The crisis of neural networks in the mid-1960s, and the closure of the perceptron controversy after Minsky and Papert's (1969) study were more important than what historical studies of (symbolic) AI usually concede. As a consequence of the closure of the perceptron controversy, symbol-processing emerged as the 'right' approach to AI and cognitive science research. One important effect of the conclusion of a controversy is the reification of the 'balance of power' emerging from that closure. As the 'winning' view institutionalises and researchers develop their activities within the accepted framework of exemplars, commitments, techniques, and institutions, it is increasingly difficult for the 'losing' side to counterbalance the established relationships of power. Because of the inertia of institutions and patterns of activity, time runs against the losing position. The

---

[57] It is interesting to see Newell's (1983) 'table of intellectual issues' of the history of AI (p. 191) and the importance which the mid-1960s have as a changing period in that table. The mid-1960s were the time of the crisis of neural networks.

in-principle plausibility of the losing side becomes increasingly impractical. B. Harvey (1981, 126) described this gap between 'principle' and 'practice' in an interesting way:

> ". . . Accepted beliefs very quickly cease to be easily comparable with rejected beliefs, because the former become the basis for future practice . . . Even when it is pointed out that the viewpoint of the losers remains logically tenable, it is difficult for the reader to remain impartial in the face of the sheer weight of numbers in the 'winning' camp."

Another aspect of reification is that the social processes (i.e. conventional decisions) at the basis of the generation and validation of scientific knowledge are 'buried' and forgotten. The result is a 'fact,' a 'black box.' Something like this happened at the end of the perceptron controversy. Minsky and Papert's (1969) arguments about single-layer perceptrons were widely interpreted as showing that the *whole* neural network approach to AI was not worth pursuing. Symbolic AI emerged as *the* approach to building intelligent machines and studying cognition computationally. The critics of neural networks were successful in *linking* their criticism of perceptrons with the 'success' of symbol-processing AI in the 1960s, and this was decisive for the closure of the controversy. This connection reduced the interpretative flexibility of the perceptron and was a mechanism of closure.

Minsky and Papert's (1969) criticism was not contested by strong enough rhetoric, and it remained as the last word in the debate. The debate was over. Some individuals continued working on neural networks throughout the 1970s and their work was important, but they were working as *individuals*, they were not powerful enough to develop a *position*. A position is, of course, a social phenomenon, comparable perhaps to a social movement. The closure of the perceptron controversy could only be revised by a mobilisation of allies and actants greater than the one carried out by Minsky and Papert (1969), something beyond the capacity of the isolated neural network researchers of the time.

It is very important to note that, although Minsky and Papert's (1969) arguments only applied to the single-layer perceptron — and even there interpretative flexibility was considerable — they were taken as showing that perceptrons in general were the wrong approach to AI. The *tone* of Minsky and Papert's (1969) criticism was very much against the whole idea of neural network research, but the link between their arguments about the limitations of the single-layer perceptron and the rejection of neural networks as a whole (including multilayer systems) was not a 'logical' or 'necessary' one at all. In section 3.3 I showed that Minsky and Papert's (1969) views on the issue of multilayer perceptrons were the weakest point of their arguments, the one most open to interpretative flexibility. It was shown then that neural network researchers of the time saw this weakness, and tried to exploit it in their response to Minsky and Papert.

Of course, the interpretation of Minsky and Papert's (1969) study as showing the uselessness of pursuing the neural network approach was a social phenomenon. Minsky and Papert represented (i.e. were the 'symbolic leaders' of) the position against neural networks in the perceptron controversy. They were key participants in the 'core set' (using H. M. Collins' term) of scientists involved in the perceptron controversy. The other position was often represented by Rosenblatt. The following remark by Harry Collins (1985, p. 148) applies therefore to Minsky and Papert:

> "For most purposes an individual's thoughts *qua individual* are of no interest. The most useful way of thinking about the goals of members of the core set is by thinking of those members as 'delegates' from the disciplines or other social and cognitive institutions which form their background."

The link between Minsky and Papert's (1969) study and the rejection of neural computing as a whole was admitted by Papert and Minsky themselves. Seymour Papert (1988, pp. 7-8) recently talked about 'universalistic attitudes:'

> "Its universalism made it almost inevitable for AI to appropriate our work as a proof that neural nets were *universally bad* . . . In fact, more than half of our book is devoted to 'properceptron' findings about some very surprising and hitherto unknown things that perceptrons can do. But in a [scientific] culture set up for *global judgement* of mechanisms, being understood can be a fate as bad as death."

Papert made this comment in 1988, when the re-emergence of neural computing (to be studied in chapters four and five) was well under way. Because of that re-emergence, Papert and Minsky are now interested in de-emphasising the association between their names and the rejection of neural networks as a whole: 'AI *appropriated* our work as a proof that neural nets were universally bad' (emphasis added). Minsky and Papert are now interested in stressing the interpretative flexibility of their 'Perceptrons' study. Another tactical comment by Papert in this respect in quotation above is that 'more than half of the book was devoted to 'properceptron' findings about some very surprising and hitherto unknown things that perceptrons could do.'

Minsky and Papert's (1969) results were indeed open to interpretative flexibility, and early neural network researchers tried unsuccessfully to exploit it. But what interests me here is that Papert admitted that their book *was* interpreted as a proof that the whole neural network approach was not worth pursuing. That was a social phenomenon, and therefore more important than Minsky and Papert's involvement in it *qua* individuals.

Minsky (Bernstein, 1981) also admitted, at least implicitly, that their 1969 book was interpreted as showing the inadequacy of the whole neural network approach. The rhetoric he used in his comments was different from that of Papert in the above quotation. This was in 1981, long before the re-emergence of neural computing.

> "In the mid-1960s, Minsky and Papert set out to kill the perceptron . . . For four years, they worked on their ideas, and in 1969 they published their book 'Perceptrons.' 'There had been several thousand papers published on Perceptrons up to 1969, but our book put a stop to those,' Minsky told me . . . The trouble was that the book was too good. We really spent one year too much on it. We finished all of the easy conjectures, and so no beginner could do anything. We didn't leave anything for *students* to do. We got too greedy. As a result, ten years went without another significant paper on the subject." (Bernstein, 1981, p. 100)

Recently Minsky and Papert have made some conciliatory comments. Papert's (1988) quotation above is an example. Minsky is reported to have 'regretted the chilling effect of his book 'Perceptrons' on neural networks' in a neural network conference in 1988 (Alternative Computers, 1989, p. 51) (I will come back to this issue briefly in section 5.3). Nevertheless, some neural network researchers like Stephen Grossberg (1989, p. 91) have not forgotten the 'bitterness' of the closure of the perceptron controversy:

> "In . . . *Perceptrons* [Minsky & Papert, 1969], Minsky and Papert focused on a single line of neural network research: Frank Rosenblatt's seminal work on perceptrons. From this analysis, they drew and actively promulgated sweeping conclusions about the entire field of neural network research, indeed about how everyone should attempt to theoretically analyze biological intelligence. It is well known that these conclusions did not favour neural network research. Everyone who managed to work on neural networks in the 1960s and 1970s can attest to the dampening effects of Minsky and Papert's anti-neural network ardor . . . I witnessed the intellectual indifference and political hostility of Minsky and Papert to these discoveries [Grossberg's early neural network contributions] when I was a professor at MIT from 1967 to 1975."

Another important element in the closure of the perceptron controversy was the 'association' between symbolic AI and the digital computer. AI research is very much an heterogeneous

network (Latour, 1987) of actants including researchers, machines (computers and other devices and systems), and software. The researchers are themselves heterogeneous: scientists, engineers, but also: psychologists, linguists, philosophers. The machines are particularly important in this heterogeneous network, and therefore it is not surprising that the 'association' (in Latour's [1987, p. 202] sense) between symbolic AI researchers and the digital computer had great importance for the closure of the perceptron controversy. In the late 1950s and early 1960s symbolic AI researchers at DARPA-funded centres like MIT, Carnegie-Mellon University, and Stanford University monopolised access to computer resources. The link between symbolic AI and the digital computer was very strong from the very beginning. The digital computer was the experimentation tool of symbolic AI researchers: a great part of research activity in symbolic AI consisted of computer simulations (using list processing programming languages like LISP).

The — by all means — impressive developments in digital computer technology from the mid-1960s onwards strengthened the position of the symbol-processing approach (for a study of these developments, see Molina [1987, chapter 2]). Analog and neural network technologies were very much on the losing side of the computer technology race. Developments in digital computer technology from the late 1960s onwards have been spectacular. For example, hardware developments include miniaturisation of electronic components (small scale integrated circuits in the mid-1960s, medium scale integration by the late 1960s, large scale integration in the 1970s, very large scale integration in the 1980s), reductions in cost per electronic component, and developments in computer power and speed (e.g. operations per second, instructions per second).

Symbolic AI grew and developed with the digital computer. The eclipse of analog computers in the mid-1960s is especially interesting here (remember the analog elements in early neural

network computers). In an analog computer, information (e.g. numbers) is represented by continuous voltage values. The history of the analog computer goes back (at least) to the 1930s, but its 'golden age' were the 1950s and early 1960s (Alternative Computers, 1989, p. 26). Analog computers were used for solving differential equations (these equations express the interplay between certain variables or forces, i.e. how one variable changes in response to changes in others). It is interesting to note that the demise of analog computers happened approximately at the same time as the crisis of neural network research. The authors of (Alternative Computers, 1989) point out at voltage precision problems in analog computers, and the accuracy and storage capacity of digital computers, and conclude (p. 27) that:

> "By about 1965, improvements in digital-computer speed and memory capacity, combined with advances in programming techniques, had made digital the technology of choice for most computer customers."

Early neural network computers, such as the Mark 1 perceptron, the Madaline, and Minos were pretty much on the 'analog side' (remember the modifiable weights), and this type of 'neural' technology was also on the losing side with the advent of digital computer technology. Early neural network researchers started to use digital computers for simulating neural networks, but the association between neural networks and the digital computer was much weaker than the one between symbolic AI and the digital computer. After all, Rosenblatt had characterised the neural network approach as opposed to the digital computer (see section 2.2). The serial character of the von Neumann computer did not favour radically parallel approaches to AI such as neural networks. And simulating neural networks in a sequential computer did not seem to many at that time the most adequate way of using the computing resources available.

Summarising, in this section I have shown that after Minsky and Papert's (1969) study the interpretative flexibility of the

perceptron was closed against the neural network position. There was nothing 'logical,' 'rational,' or 'natural' in the link from Minsky and Papert's arguments to the rejection of neural network research as a whole. However, that link was made, and the closure of the controversy — a social phenomenon — crystalised after Minsky and Papert's (1969) study. Neural network researchers were not powerful enough to contest the interpretation of Minsky and Papert's (1969) study as showing that the neural network approach (as a whole) was not worth pursuing. In addition, the anti-neural network position was successful in mobilising powerful allies in its favour. The most powerful ally (and closing mechanism) in this respect was symbol-processing AI. Another important association was the one between the anti-neural network position and the digital computer. However, these associations were not 'logical,' but social. The link between the limitations of the single-layer perceptron and the adoption of the symbol-processing approach to AI — a socially constructed link — was a decisive factor in the closure of the perceptron controversy. I chapters four and five I show that it took a long time for neural network researchers to break that link.

◆ FOUR

New Connectionism

4.1 Parallel distributed processing

In this section I study certain aspects of the re-emergence of neural network research in the 1980s. After making some comments about neural network research in the 1970s, I look at some developments of the early 1980s such as the situation of symbol-processing AI and developments in computer (and parallel computer) technology. I approach the process of re-emergence of neural network research in the 1980s as a process of formation and development of an heterogeneous network where diverse allies and actants were enroled. The Parallel Distributed Processing (PDP) group played a key role in that process. I look here only at some aspects of that enrolment process. In later sections I will examine some of the most important innovations developed in the 1980s.

In this section I show that, by linking several developments in the early and mid-1980s, the PDP group brought neural network research back to the AI-cognitive science arena. In particular, the PDP researchers linked (and exploited for their purposes) the following elements: neural network research carried out in the 1970s, the problems that symbolic AI was having in studying certain cognitive capabilities, parallel computing, and the similarities between symbolic AI and neural networks in certain areas (e.g. semantic networks).

After the closure of the perceptron controversy activity in neural networks decreased to its minimum levels. Throughout the 1970s a small number of researchers did carry out research in neural networks and related topics, but they were far from the main centres of activity in AI and cognitive science research, where symbol-processing continued to be the dominant approach over the years. This was much more so in the United States than in Europe, where activity related to neural networks (but more

oriented towards neurobiology and psychology rather than to AI-like research) remained relatively strong. Researchers who worked in neural networks and related topics in the 1970s include Christoph von der Malsburg, David Willshaw, Teuvo Kohonen, Geoffrey Hinton, and Igor Aleksander in Europe, Michael Arbib, Stephen Grossberg, James Anderson, Jack Cowan, and Leon Cooper in the United States, and K. Fukushima and S. I. Amari in Japan. The list is not exhaustive, but it can be seen that the relative strength of Europe is much more significant than in the early neural network period.

The importance of neural network-like research in Europe was reflected in Sir James Lighthill's (1973) important report on the state of AI in the early 1970s for the UK Science Research Council (SRC). One of the three main areas of AI research studied by Lighthill was (neuroscience and psychology-oriented) computer-based central nervous system research in man and animals (the other basic categories in the Lighthill report were symbolic AI/advanced automation and robotics). Lighthill (ibid., pp. 19-21) concluded that success of work under the category 'computer-based central nervous system research' would depend on its close relationships with psychology and neurobiology, in the same way as work on advanced automation/symbolic AI would depend on its close association with its application area (engineering). He also concluded that robotics research should integrate in those areas.[58] What is interesting about the Lighthill report is that neural network-like research received as much attention as symbolic AI as a subarea of AI research. This shows the relative strength of (neuroscience-oriented) neural network research in the UK, as compared to the situation in the US where the symbolic approach was much more clearly dominant.

Analysing neural network-like research in the 1970s is out of the scope of this dissertation. One of the most important

---

[58] Fleck (1982, p. 205) interpreted this as a denial of an autonomous realm for AI research.

characteristics of this period is the retreat of neural networks from AI research to more neuroscience-oriented and psychology-oriented research areas. An example of neuroscience-oriented neural network research is von der Malsburg and Willshaw's work on topographic maps of neural connections. Building upon Christoph von der Malsburg's (1973) work on self-organisation and topographic mapping using neural networks, Willshaw and von der Malsburg developed a two-sheet network (with its learning algorithm) which is known as the 'tea-trade' model (because of the analogy they used to describe their model intuitively) (von der Malsburg & Willshaw, 1977; Willshaw & von der Malsburg, 1979). In the early 1980s, inspired by Willshaw and von der Malsburg's idea of computational topographic mapping, Kohonen (1982, 1984) developed an algorithm for unsupervised neural networks.

More psychology-oriented neural network research in the 1970s includes work in content-addressable associative memory, carried out by researchers including Anderson, Kohonen, and Willshaw. The origins of associative memory neural network research go back to the time of early neural networks, to work carried out by researchers like W. K. Taylor (1956, 1959) in University College London and Karl Steinbuch (1961) in Germany (see: Cowan & Sharp, 1988, pp. 92-94; Aleksander & Morton, 1990, pp. 57-58). Another important development of the 1970s was Stephen Grossberg's (1976a, 1976b, 1976c, 1978) work on unsupervised neural networks. Many of the most important neural network contributions of the 1970s were reprinted by Anderson and Rosenfeld (1988).

It is also out of the scope of this dissertation to look at developments in symbolic AI in the 1970s. Although the symbol-processing approach dominated in AI throughout the 1970s, controversies and changes in funding patterns for symbolic AI were not rare in the United States and in Europe. Examples of this are, as Fleck (1982, p. 192) pointed out, cuts in funding for AI (mainly robotics) in the UK after the Lighthill report,

subsequent cuts in robotics in the US, and ARPA's insistence on mission-oriented AI research. The cuts for robotics research in the US affected the 'Shakey' mobile robot project of the formerly neural network group at SRI (see section 3.2). In the 1970s, the overall trends of funding for science in the US were much more applications-oriented (Dickson, 1988).

Towards the early 1980s, various new developments started to alter the situation in AI and related disciplines. I will comment briefly on some of these developments now (a detailed discussion of these issues is out of the scope of this section).

In the early 1980s symbolic AI went from a stage of 'establishment' and institutionalisation to one of greater growth and commercialisation (Fleck, 1987). This new phase was triggered by the Japanese Fifth Generation Project. In October 1981, the Ministry of International Trade and Industry of Japan (MITI) launched a ten-year, $850 million computer technology project, in which they emphasised the importance of AI (especially natural language and knowledge-based information processing). The US and UK governments reacted quickly by launching their own computer technology programmes (namely Microelectronics and Computer Technology Corporation [MCC], DARPA's Strategic Computing programme, and the Strategic Defense Initiative [SDI] in the United States, and the Alvey programme in Britain). This climate favoured AI research.

The early 1980s were the time of the most important commercialisation of symbolic AI so far: expert systems. Basically, expert systems are composed of a knowledge-base (where knowledge relevant for a certain domain is represented) and techniques for making inferences from that base in a particular situation or problem. In these knowledge-based information processing systems emphasis if laid on (symbolic) representation and on the ability of the computer to carry out structure-sensitive transformations of those representations. Expert systems have been applied to a great variety of situations. However, symbolic AI research has not been so

successful in other areas such as speech recognition, pattern recognition, and common-sense and heterogeneous reasoning. It will be seen below that neural network researchers took advantage of these weaker points in their rhetoric in favour of connectionism in the 1980s.

Computer technology developments were an important aspect of the early 1980s. Hardware developments included miniaturisation, increases in computing power, and reduction in cots. The early 1980s were the time of very large scale integration (VLSI), i.e. the development of single chips with hundreds of thousands of components on them.

At around the same time, the limitations of the von Neumann computer architecture were becoming increasingly apparent (Peláez, 1988). The separation between memory and central processing unit (linked by a 'connecting tube') in a von Neumann computer imposes a sequential (one operation at a time) style of computation. One obvious limitation of this style of computing is speed. By the early 1980s, several approaches to parallel computing (the use of more than one processor working concurrently in a problem) were emerging.[59] The cost of microprocessors had decreased significantly by then, but parallelism involves many issues which are not well understood or developed yet. A good example of this, pointed out by Peláez (1988), is the question of software for parallel computers. The problem of parallelism can be seen as the question of "how problem-solving can be distributed across a network of interacting , concurrently active processors" (Arbib 1989, p. 186). This question is being developed from many different points of view including computer architectures (see previous

---

[59] According to 'granularity,' parallel architectures can be 'coarse grain' (small number of sophisticated processors), or 'fine grain' (large number of simpler processors). According to the instructions received by each processor, they can be single instruction/multiple data (SIMD) or multiple instruction/multiple data (MIMD, with each processor receiving its own instructions). For a history of parallelism and supercomputing, see Hockney & Jesshope (1988, pp. 2-53).

footnote), distributed AI, computer networking, parallelism in machine vision systems, and neural networks.[60]

These trends towards parallelism in the 1980s favoured the resurgence of interest neural networks as a (radical) parallel computer architecture. (Of course, they do not explain the neural network innovations of the 1980s.) One of the radical aspects of parallelism in neural networks is that computation is defined at the subsymbolic level, and symbolic entities are seen as properties emerging from the parallel interaction of many simple processing units (simplified neurons). Neural network parallelism is massive and brain-like, very different from other parallel architectures.

In the early 1980s, neural network researchers argued that the information-processing power of the brain comes from its parallelism. Feldman and Ballard (1982) formulated what is sometimes called the '100 step constraint.' This constraint is an approximate measure of the time required by the human brain to carry out certain complex cognitive processes such as, for example, recognising a human face. Within the neural computing community the 100 step constraint is seen as a strong argument in favour of neural network-like parallelism.

> "Neurons whose basic computational speed is a few milliseconds must be made to account for complex behaviors which are carried out in a few hundred milliseconds (Posner, 1978). This means that entire complex behaviors are carried out in less than a hundred time steps. Current AI and simulation programs require millions of time steps . . . The firing frequencies of neurons range from a few to a few hundred impulses per second. In the 1/10 second needed for basic mental events, there can only be a limited amount of information encoded in frequencies." (Feldman & Ballard, 1982, pp. 484 and 487)

---

[60] Arbib (1989, p. 187) pointed out that 'cooperative computation' (heterogeneous networks of special-purpose and general purpose subsystems), 'perceptual robotics,' and 'learning' may be the main characteristics of Sixth Generation Computing.

With the advent of parallel computers and supercomputers, researchers started to compare computer power and brain-style information processing. Speculations were made about the number of operations per second in the brain as compared with the most powerful computers. Sejnowski (1987, pp. 206-207) estimated the minimal amount of digital computation necessary to simulate neural operations in real time in 10^{15} operations per second, about 10^5 times greater than the largest general purpose digital computer, and concluded that:

> "The cost of computing has decreased by a factor of about 10 every 5 years over the last 35 years . . . If this continues, then it will take about 25 more years (2015) before processing power comparable ti that in the brain can be purchased for $3 Million . . . It is very unlikely, however, that this goal can be achieved with the current technology: new technologies, perhaps based on optical computing, are needed."

Computing power in abstract is important because it allows increasingly more powerful simulations of (i.e. experimentation with) neural networks, but of course many other problems remain in neural network research (e.g. architecture and organisation of the network, learning). In section 3.4 it was seen that symbolic AI developed a strong association with the increasingly successful von Neumann (serial) computer technology. This association did not favour neural networks, which are naturally parallel. Recent developments in parallel computing are more favourable for neural computing. The 'association' between computer technology and neural network research is more powerful this time round, although neural network researchers, with their radical, subsymbolic, and massive parallelism approach, have to compete with other approaches to and uses of parallelism.

One subarea of symbolic AI where parallelism has often been used is machine vision. In the 1970s machine vision was very much within the symbolic AI umbrella (although using their own

particular techniques). David Marr's change in the early 1980s from neural network research (Marr,1969; 1970; 1971) to machine vision within symbolic AI was significant in this respect, and reflects the closure of the perceptron controversy.

> "There seemed no reason why the reductionist approach could not be taken all the way. I was myself caught up in this excitement. . . [But] in the early 1970s it gradually became clear that something important was missing that was not present in either of the disciplines of neurophysiology or psychophysics . . . [Now] gone is any explanation *in terms* of neurons — except as a way of implementing a method . . . The message [in the 1970s] was plain. There must exist an additional level of understanding at which the character of the information-processing tasks carried out during perception are analyzed and understood in a way that is independent of the particular mechanisms and structures that implement them in our heads." (Marr, 1982, pp. 14-15, and 18-19)

Vision researcher Harry Barrow (1989, p. 12) admitted that this methodology was sometimes espoused to extremes by machine vision researchers, and concluded that:

> "The effects of the third (hardware implementation) and second (representation and algorithm) levels may, in fact, make themselves felt even at the first level and should affect assumptions and decisions there."

Vision research notions such as the parallel interaction between many local features in the interpretation of an image were an area where symbolic AI and neural networks were closer than usual (Ballard et al., 1983). Hinton and his colleagues were motivated by this type of problems when they developed their Boltzmann machine network (to be studied in section 4.2). Machine vision research was a potential ally of neural network research. But the methodological extremes mentioned by Barrow above (i.e. the neglect for the implementation level) had to be overcome first.

Thus there were several heterogeneous allies that neural network researchers could try to enrol. (And I have not mentioned all of them. An obvious ally which has not been mentioned were the researchers who had worked in neural networks throughout the 1970s.) Latour (1987) used the notion of 'heterogeneous network' (where researchers, machines, and a variety of objects and factors are intertwined) mainly in microsociological (laboratory level) case studies (although these laboratories always have an 'outside'). Nevertheless, one could think of the emergence of neural network research in the 1980s as the development of a big 'heterogeneous network.' Then, the reopening of the perceptron controversy can be seen as a 'trial of strength,' a process of enlisting of (heterogeneous) allies and resources through which neural network researchers were finally able to contest Minsky and Papert's (1969) 'black box.'

In the rest of this section I look at some 'contextual' aspects of this 'enlisting' process. I discuss these aspects before I look at neural network innovations only for tactical purposes: these aspects can be a sort of introduction to the innovations that I study in the coming sections. Discussing contextual and organisational aspects first does not mean, therefore, that these factors 'cause' scientific innovation (more about this in section 4.2).

The Parallel Distributed Processing (PDP) group played an important role in the above mentioned 'enlisting' process. Their scientific and 'lobbying' (so to speak) activity was crucial for bringing neural network research back to the AI and cognitive science arena. Several of the most important innovations in neural computing in the 1980s, such as the Boltzmann Machine network and the back-propagation network, were developed by researchers belonging to the PDP group. I will look at these innovations in sections 4.2 and 5.2. Here I concentrate on the history of the PDP group and on the context in which it developed.

Two events help locate the origin and the apogee of the PDP group: the 1979 San Diego meeting and the publication of the PDP volumes in 1986 (Rumelhart, McClelland, & the PDP Research Group, 1986; McClelland, Rumelhart, & the PDP Research Group, 1986) respectively.[61] The 1979 La Jolla conference was organised by Geoffrey Hinton and James Anderson. Later Hinton and Anderson (1981) published a collection with the papers presented at that meeting and a few other contributions.[62] Anderson (in the US) and Hinton (in the UK) had worked in neural networks in the 1970s. Anderson (1972) had worked on associative memory, and Hinton (1977) on co-operative processing in vision under H. C. Longuet-Higgins' supervision.[63]

According to James Anderson (interview) the purpose of the La Jolla conference was to do something small and informal, where researchers who had been working on similar topics without knowing well each other's work could get together and start some kind of useful interaction:

> "We felt that people were really doing the same thing, but they didn't know it. There were people that were doing similar things but they weren't in communication. I think that, overall, that was a correct analysis of the time."

David Rumelhart, who later became a leading member of the PDP group sees the motivation for the La Jolla meeting and the PDP group in similar terms:

---

61 The members of the PDP group were Chisato Asanuma (Salk Institute), Francis H. C. Crick (Sal Institute), Jeffrey Elman (Linguistics-UCSD), G. E. Hinton (then at Computer Science-Carnegie Mellon Univ., now at Comp.Sci.-Univ. of Toronto), Michael I. Jordan (Computer & Inf. Sci.-Univ. of Mass.Amherst), Alan H. Kawamoto (Psychology-Carnegie Mellon Univ.), J. L. McClelland (Psychology-Carnegie Mellon Univ.), Paul W. Munro (Inf.Sci.-Univ.of Pittsburgh), D. A. Norman (Cog.Sci.-UCSD), Daniel E. Rabin (Intellicorp-Mountain View CA), D. E. Rumelhart (then at Cog.Sci.-UCSD, now at Stanford Univ.-Psychology), T. J. Sejnowski (Salk Institute), P. Smolensky (Comp.Sci.-Univ.of Colorado Boulder), Gregory O. Stone (Math.-Boston Univ.), R. J. Williams (then at Cog.Sci.-UCSD, now Northeastern Univ.Comp.Sci.Boston), and David Zipser (Cog.Sci.-UCSD).

62 Reprinted as (Hinton & Anderson, 1989).

63 Longuet-Higgins developed the analogy between the hologram (and other devices such as the correlograph) and associative memory, and supervised D. Willshaw and G. Hinton. See: Willshaw, Buneman, and Longuet-Higgins (1969).

> "Fukushima, Grossberg, Amari, Arbib, Anderson, Kohonen . . . There were these people that were working, but they weren't really working together, they were pretty much in isolation, and pretty much not very visible . . . By and large, the neural net people were distributed, not working together, and overall not very much seen by the AI-psychology-engineering-neuroscience community." (Rumelhart, interview)

The PDP group was formed in 1981 at the Institute for Cognitive Science of the University of California-San Diego (UCSD, La Jolla, California) (Rumelhart, interview). Apart from the PDP meetings there were also several other small meetings, held approximately once a year from 1981 to 1984 (Feldman, interview), in which PDP researchers and a few other researchers like Jerome Feldman participated.[64] These meetings seem to have been important for the formation of what could be called the 'new connectionist movement.' Feldman (interview) described them in the following terms:

> "These were very tough but intellectually supportive meetings. Most of the people who were there would view those as some of the most fruitful scientific meetings they have ever been in. There was a lot of fun, it was very productive. To the extent that there was a kind of connectionist movement, I think it took place in those meetings, not so much in the 1979 San Diego meeting, which sort of brought a bunch of people together, but the four or so meetings after that. There was really quite a small kind of revolutionary cell in the early days of connectionist stuff. There was a first explosion, we had an special issue of 'Cognitive Science' [Cognitive Science,

---

[64] Feldman was at the University of Rochester at the time, and currently he is also director of the International Computer Science Institute (Berkeley, California). He was also at the 1979 la Jolla meeting. A 'traditional' computer scientist by formation, Feldman (Feldman & Ballard, 1982) developed his own particular approach to neural networks, which is sometimes called 'structured neural networks.' "My background was very traditional computer science . . . In fact, the San Diego people were actually a little reluctant to invite me to the 1979 meeting, because they knew about my previous work as a very traditional computer scientist" (Feldman, interview). Feldman et al. (1988) is a recent statement of Feldman's structured neural network approach. Feldman often uses symbolic (or localist) representations in his systems. Feldman's approach to representation is discussed in (Feldman, 1988).

1985]. They asked us to make an special issue. It was obviously a recently hot topic by whenever that was."

Feldman's words reflect a 'political willingness' (so to speak) of the participants in those meetings to develop some sort of 'connectionist movement,' the small group meeting throughout those years being the 'revolutionary cell' or 'vanguard' of the movement.

The rhetoric used by the PDP group in defending and legitimising their position was similar to that used by Frank Rosenblatt (see section 2.2) in the late 1950s and early 1960s. They claimed that there were important aspects of human cognition which could not be explained or model using a von Neumann computer-based approach. In other words, they tried to focus the debate on those issues where the symbol-processing approach to AI was weakest and could not deliver all the promised results. An important point here is the alliance between researchers who had been working on neural networks throughout the 1970s (e.g. G. Hinton and J. Anderson) and those who had been working within the symbol-processing approach (e.g. Rumelhart, McClelland, and Norman).[65] Researchers like Rumelhart or Norman were interested in an alternative computational framework within which they could study and model 'soft cognition,' i.e. capabilities such as the flexibility of common sense reasoning, resistance to error, and distributed memory. Norman's (1986, p. 537) words can be interpreted in this sense:

> "Some years ago, Bobrow and I listed some properties we felt were essential components of the human cognitive system (Bobrow & Norman, 1975; Norman & Bobrow, 1975, 1976, 1979). We constructed our lists through observations of human behavior and reflection upon the sort of

---

65 For example, Rumelhart had been working on the problem of frames in symbolic cognitive science: ". . . The concept of *schema* or related concepts such as scripts, frames, and so on. These large data structures have been posited as playing critical roles in the interpretation of input data, the guiding of action, and the storage of knowledge in memory . . . It was struggling with the concept of schema and some of its difficulties that led one of us (David Rumelhart) to an exploration of PDP models to begin with" (Rumelhart, Smolensky, McClelland, & Hinton, 1986, p. 7).

> processing structures that would be required to yield that behavior. We concluded that the system must be robust, relatively insensitive to missing or erroneous data and to damage of its parts. Human cognition appears to work well in the face of ambiguity, incompleteness, and false information. On the one hand, the system is not only robust, it is also flexible and creative. On the other hand, the system continually makes errors. Speech is riddled with incomplete sentences, with erroneous words and false starts. Actions are riddled with 'slips' and mistakes. People have learned to deal with these problems, in part by elaborate (although mostly subconscious) correcting mechanisms in language, in part through humor and tolerance for the slips that characterize everyday behavior. Nonetheless, these characteristics of human cognition seem to provide important clues about the nature of the processing mechanisms. We argued that the system had to work by descriptions rather than precise specifications, by partial information rather than complete information, and by competition among competing interpretations. We argued for a set of essential properties: graceful degradation of performance, content addressable storage, continually available output, and an iterative retrieval process that worked by description rather than by more traditional search."

The alliance between researchers who had been working on neural networks in the 1970s and researchers coming from the symbol-processing tradition but who were critical about it was very important because it helped bring neural network research back to the AI-cognitive science arena. But researchers in favour of the symbol-processing approach were not (and are not) ready to abandon the terrain that connectionists claim. The common-sense reasoning (or nondemostrative inference) 'front' is an example, as this statement by Jerry Fodor and Zenon Pylyshyn (1988, p. 30) shows:

> "Classical theory construction [i.e. the symbolic approach] rests on the hope that syntactic analogues can be constructed for nondemostrative inferences (or informal, commonsense reasoning) in something like the way that proof theory has provided syntactic analogues for validity."

This is a 'bear with us, we will deliver results' tactic, and shows that symbol-processing researchers admit (at least indirectly) that they are having trouble modelling certain cognitive capabilities. A tactic which researchers in favour of connectionism used in responding to this type of 'bear with us' argument was to emphasise the relative importance of common-sense reasoning within perception cognition.

> "Common-sense defeasible inferences, if derivable at all on a proof-theoretic account, must be extremely complex. For people, these inferences are just *common sense*. This mode of inference underlines comprehension, categorisation, perception, and action . . . It is the basis of all cognitive performance." (Oaksford, Chater, & Stenning, 1990, p. 83)[66]

Researchers who had worked within the symbol-processing approach but remained critical, such as Norman, appeared optimistic about the results of their alliance with researchers who had worked in neural networks in the 1970s. They argued that they now have an alternative computational framework within which they can study and model cognition. This is seen in Norman's (1986, p. 537) remarks about the work carried out by the PDP group:

> "Although [those] requirements [see quotation by Norman above] seemed to us both necessary and sensible, supported by the evidence, a common complaint was that these were hand-waving speculations, dreams, that there was no method for implementing these concepts. And if they couldn't be built, then they didn't exist — the ideas were wrong. Well, the PDP mechanisms described in this book [Rumelhart, McClelland, & the PDP Research Group, 1986; McClelland, Rumelhart, & the PDP Research Group, 1986]

---

[66] Chater and Oaksford (1990, pp. 102-103) give the following example of common-sense inference: "Just about any everyday generalisation succumbs to indefinitely many counter-examples. If I see Fred going past my window at 9.00 a.m., I know he's about to buy his morning paper. But not if it's Christmas day, since there are no papers; not if he's being mugged; not if he's already reading *The Times*. These possibilities override our generalisation that Fred buys a paper just after passing my window every morning."

have exactly the properties we required. Hurrah for our side."

Oaksford, Chater, and Stenning (1990, p. 77) made a similar claim:

> "Connectionism may provide the competition which promotes the mismatch between the Classical model [i.e. the symbolic approach] and the empirical data to falsificatory status."

The interest of researchers who were studying cognitive behaviour topics such as perception, memory, object categorisation, natural language, and common-sense reasoning in an alternative to the 'formal,' 'deductive' symbolic AI and cognitive science approach in apparent in these quotations.

But let me come back to the early years of the PDP group. The PDP enterprise was based on to the association of several elements in a kind of heterogeneous network. The elements that the PDP researchers linked to each other throughout the years were apparent in the contributions of the 1979 La Jolla meeting. One linkage which the PDP researchers were particularly interested in making was the one between distributed (neural network-like) memory and parallel computing. Anderson and Hinton (1981, p. 11) made this clear:

> "Von Neumann machines are based on the idea of a sequential central processor operating on the contents of a passive memory in which data-structures simply wait around to be inspected or manipulated. This conception of memory is shared by most psychologists and is embodied in the spatial metaphors we use for talking about the process of remembering . . . The memory models presented in this volume assume a very different basic architecture. Instead of a sequential central processor and a passive memory there is a large set of interconnected, relatively simple processors, which interact with one another in parallel via their own specific hardware connections. Changes in the contents of memory are made by forming new connections or changing the strengths of existing ones. This overcomes a major bottleneck in von Neumann machines, which is that

> data-structures or programs in memory can only have effects via the sequential central processor, so that it is impossible to mobilize a large quantity of knowledge simultaneously."

One sentence of this quotation is particularly interesting for my purposes here: 'changes in the contents of memory are made by forming new connections *or* by changing the strengths of existing ones' (emphasis added). The expression 'forming new connections' here refers to the 'semantic network' tradition. Semantic networks are graphic-like structures composed of nodes and links. Nodes represent concepts, properties, or names, and links represent relationships between them. One important property of these networks is 'inheritance:' a concept will be linked to instances of it, and therefore if the node corresponding to the former is activated it will activate all the instances.

The differences between semantic networks and neural networks are considerable. In a semantic network nodes represent properties at the symbolic level, e.g. 'Clyde' and 'elephant' (this is sometimes called 'localist representation'). If the node standing for 'elephant' malfunctions (and this would not be very rare in a network with many nodes) the whole concept of 'elephant' is lost. Links represent relationships such as 'is a.' Links are hardwired, that is each time one wants to create a relationship between two concepts, or between a concept and an object, the corresponding link has to be created, and that link (which can be implemented using a simple processing unit) can only be used for that purpose.

One important early work in semantic networks was that of Ross Quillian (1968), a student of Marvin Minsky. One aspect of Quillian's semantic network which is particularly interesting here is that Quillian introduced variability in the relationships represented in his network. The relationship between a token node (i.e. an object or name) and a type node (i.e. a property or concept) could have nine degrees of intensity in Quillian's

network (ibid., p. 231). This specification was included in the token node, not in the connecting link.

Several researchers who participated in the 1979 La Jolla conference, such as Scott Fahlman(1981) and Jerome Feldman (1981) were working pretty much in the 'semantic network' tradition. The biggest contrast between this work and later work by the PDP work is the use of localist (non-distributed) representations. In the early 1980s Rumelhart and McClelland also used localist representations (i.e. units representing letters or words) in their early research on word recognition (their word perception or interactive activation model) (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982). The use of semantic networks and localist representations in the early stages of the evolution of the PDP group can be seen as reflecting the 'transition' from the symbolic approach to neural networks. Later PDP researchers emphasised much more the importance of distributed representations and learning.[67]

Thus semantic networks were somehow linked to neural network research in the early stages of the PDP group.[68] In a sense, a semantic network was the most similar thing to a neural network that one could find within the broad umbrella of symbolic AI. Nevertheless, it is important to emphasise that, in the late 1960s, semantic network research was seen as opposed to neural network research. In the 1968-1969 period Minsky published two important books: 'Semantic Information Processing' (Minsky, 1968b) and 'Perceptrons' (Minsky & Papert, 1969). The former showed 'the way to go,' so to speak, and the latter 'the way not to go.' In 'Semantic Information Processing,' the semantic network approach was represented by the paper by Quillian (1968) that I mentioned earlier. It is interesting to note that in Minsky's (1968, p. 6) view semantic network research was *opposed* to neural network research.

---

[67] Feldman did not. As it was said earlier, he continued to use 'localist representations' and his own, more structured, approach.

[68] For a more recent paper in which localist and distributed models are presented within the same broad framework, see Fahlman and Hinton (1987).

> "Quillian [1968] . . . builds a network of objects and relations . . . and this is able to compare the plausibilities of different interpretations by the strengths of the chains linking the various pairs of meanings . . . In this development one can see the return to life of some of the basically sound concepts of association psychology that became frozen caricatures in a generation of premature linear and stochastic learning theories."

Rumelhart himself (interview) confirmed this opposition:

> "[In the late 60s and early 70s] I actually moved more into the AI area, spent most of my time in the AI applications to psychology, what AI had to teach. I was quite intrigued by the work of Quillian. In fact, I read both of those books [Minsky's (1968) 'Semantic Information Processing,' and Minsky and Papert's (1969) 'Perceptrons'] at the same time. I taught courses in both of them in fact, and my own work followed up on the semantic information processing end, rather than on the perceptron end."

Thus the links between semantic networks and neural networks had not been made at all in the late 1960s. They were made later by the PDP group. It is interesting that while in the 1960s semantic networks were seen as an alternative to neural networks, in the late 1970s and early 1980s their similarities were exploited by the PDP group.

But the PDP researchers did not stop in semantic networks. They used some elements of semantic networks in their early systems, but they were aiming at a more radical 'move.' Hinton (1981, pp. 161-162) argued that there was a 'radically' new way of implementing 'semantic networks:'

> "There are two very different ways of implementing semantic networks in networks of simple hardware units. The obvious approach is to make different nodes in the semantic net [concepts, representations] correspond to different hardware units and to make links between semantic nodes correspond to hardware links between units . . . A *radically different* approach is to make each node in the semantic net correspond to a particular pattern of

> activity on a large assembly of units. Different semantic nodes may then be represented by different patterns of activity on the same set of units . . . The semantic net formalism can then be seen as a crude description of the interactions between complex patterns of activity . . . The interactions between concepts that are formalized as a single link in a semantic net are actually generated by millions of simultaneous interactions at the level of their microstructures." (emphasis added)

The 'radically different implementation of semantic networks' was simply neural networks. But it is interesting to note that Hinton defined his 'radical alternative' using the traditional (*status quo*) terminology of semantic networks. He was making the move from the most neural network-like type of symbol-processing to actual neural networks.

Another similar 'move' by Anderson and Hinton in the early stages of the PDP group was to argue that a neural network system composed of many simple interacting units could be described at two levels. One was the usual level in neural network research, that is the level of the interaction between processing units. The other one was a 'higher' or more 'abstract' level. At this second level, the behaviour of a neural network system could be described as symbol-transformation (the equivalent of a symbol being a pattern of activation of many interacting units). So, after all, connectionist systems were somehow related to symbol-processing. They could be described in terms of the symbolic approach.[69] Another effect of this type of move was to emphasise the relevance of neural networks for the study and modelling of cognition and intelligence (to bring neural network research back to the AI-cognitive science arena).

> "A symbol, for example, could be a pattern of activity in a large group of hardware units. Provided this pattern is reproducible and regularly causes other such patterns, it is possible to implement symbol processing by the interactions of these patterns . . . At the high level,

---

[69] PDP researcher Paul Smolensky's (1988) recent move to call neural network research the 'subsymbolic' paradigm could be seen as a similar move.

> reproducible patterns of activity can be denoted by abstract symbols, and regular interactions between them can be captured by explicit rules. This kind of a description can be implemented rather directly on a conventional digital computer." (Anderson & Hinton, 1981, p. 30)

In sections 4.2 and 5.2 I will study the important innovations developed by members of the PDP group. In this section it has been seen that they also played an important role in linking several developments of the early 1980s together. In particular, they linked neural networks with the cognition and intelligence issues at which the symbol-processing approach was weaker, and in doing so they brought neural networks — and the work that had been done during the 1970s by researchers like Hinton, Anderson, Kohonen, von der Malsburg, and Grossberg — back to the AI-cognitive science arena. They also claimed that neural network systems were massive parallel, and therefore neural network research could be linked to developments in parallel computing technology.

The result of the research carried out by the PDP group — the so-called 'PDP books' (Rumelhart, McClelland, & the PDP Research Group, 1986; McClelland, Rumelhart, & the PDP Research Group, 1986) — were a major development in the emergence of neural networks in the second half of the 1980s. Minsky and Papert (1988, p. 247) called the PDP books the 'connectionist manifesto.' I will come to the innovations contained in those books in later sections. Now I would like to conclude this section by emphasising the PDP group's overt use of a 'here is an alternative AI approach' tactic in (Rumelhart, McClelland, & the PDP Research Group, 1986, pp. 3-146). This may seem of little importance, but the clear and explicit formulation of neural network research as an alternative approach to AI and cognitive science contained in those pages was the introduction to neural networks for most of the new comers to the field (and therefore most of the people in the field; see section 5.3).

It seems that the PDP researchers were also important in 'enrolling' in the neural network enterprise powerful allies such

as DARPA or the National Science Foundation (NSF). Forsyth (1989, p. 25) reported that:

> "In 1986, a team headed by Rumelhart and McClelland at the University of California, San Diego, submitted a report to DARPA (and later to its civilian equivalent NSF . . .). In it they argued that research into Parallel Distributed Processing (PDP) [i.e. neural networks] had been seriously underfunded for a decade at least. Basically they advocated a switch of resources into the PDP field — their name for connectionism. This report received a favourable reception, so much so that both DARPA and NSF have recently persuaded some quite high-powered institutions (including the AI group at UCLA) to alter the whole thrust of their work."

But DARPA's involvement in neural networks (to be studied in section 5.3) came after important innovations by the PDP group and other researchers had been developed. In the coming sections I will look at some of the most important of those innovations, and at the reopening of the perceptron controversy in the second half of the 1980s.

In this section I have studied certain aspects of the re-emergence of neural network research in the 1980s. In particular, I have looked at the PDP group's role in bringing neural network research back to the AI arena. I have showed that the PDP researchers played an important role in 'connecting' neural network research with certain favourable factors and developments of the early and mid-1980s. These developments and factors include the neural network researchers of the 1970s, researchers who were having problems in studying certain cognitive capabilities within the symbol-processing approach, and trends towards parallel computing. I have also showed that certain parts of the work carried out by the PDP researchers (such as the semantic networks issue) reflects the transition from the symbol-processing to the neural network approach. In the coming sections I will look at the innovations developed by PDP researchers and others in the 1980s.

4.2 Networks with symmetric connections: metaphors and innovation in neural computing

In this section I analyse two very important neural network innovations carried out in the early and mid-1980s, namely John Hopfield's (1982) network and David Ackley, Geoffrey Hinton, and Terrence Sejnowski's (1985) 'Boltzmann machine' network. I show that a 'metaphor scheme' (Barnes, 1974) is useful for the study of the development of these innovations. In addition, I discuss the importance of the Boltzmann machine work (a solution to the problem of training multilayer systems) as an antecedent of the reopening of the neural network controversy. I conclude by examining the similarities between the Boltzmann machine and certain developments from the cybernetics movement of the 1940s and 1950s.

Hopfield's and Hinton and colleagues' work provoked the interest of physicists in neural computing. It is sometimes assumed that the arrival of physicists to neural computing brought about the acceptance of the field. Some aspects of the growth of the neural network research community will be discussed in section 5.3. In this section (4.2) I show that, even though the migration of physicists and other researchers to neural networks and the subsequent reorganisation of the structure of the neural network community in the late 1980s showed the acceptance of neural computing as a legitimate area of research, that migration and the subsequent reorganisation were consequences, and not causes, of the innovations by Hopfield, Hinton and colleagues, and others, and of the reopening of the neural network controversy. Of course these innovations (and others to be studied later) did not come in a vacuum. They were part of a process of enrolment of actants, allies, and resources which finally caused the reopening of the perceptron controversy and the subsequent

migration of physicists, computer scientists, and engineers to the neural network field. It is very important to separate clearly the enrolment process from the subsequent growth and reorganisation of the neural network research community. Innovations like the ones to be discussed in this section and the one analysed in section 5.2 belong to the enrolment process, and therefore they were not a consequence of the migration of many physicists or other 'respectable' scientists and engineers to neural networks.

In this section I show that the innovations developed by both Hopfield and Hinton and colleagues can be understood within a 'metaphor scheme' (Barnes, 1974). Ackley, Hinton and Sejnowski's (1985) Boltzmann machine system was particularly important because they developed, for the first time, a powerful technique for training multilayer neural networks. Towards the end of the section I discuss the similarities between these developments in the 1980s and some antecedents of the use of statistical physics as a metaphor in information processing within the cybernetics movement of the 1940s and 1950s.

The following statement from the journal 'Nature' shows the importance that is sometimes given to Hopfield's work in today's neural network community and surroundings.

> "Neural networks are mostly elaborations of a crisp formulation by J. J. Hopfield (1982)." (Maddox, 1987, p. 571)

It seems that some people even think that Hopfield has 'invented' neural networks.[70] Some neural network researchers, such as James Anderson, have emphasised the importance of Hopfield and his work for the legitimisation of neural computing in the 1980s.

> "John Hopfield is a distinguished physicist. When he talks, people listen. Theory in his hands becomes respectable. Neural networks became instantly legitimate, whereas before, most developments in networks had been the

[70] Hecht-Nielssen (1990, p. 19) warned against this belief.

province of somewhat suspect psychologists and neurobiologists, or by those removed from the hot centers of scientific activity" (Anderson & Rosenfeld, 1988, p. 457)

"John Hopfield was a well-known physicist with important connections at CalTech and Bell Labs. His interest and work in neural networks legitimized the field within the physics community" (Anderson & Hinton, 1989b, p. 2)[71]

R. Hecht-Nielsen (1990, p. 19) emphasised the importance of Hopfield's work in his recent neural computing textbook:

". . . By the beginning of 1986, approximately one-third of the people in [neural computing] . . . had been brought in directly by Hopfield or by one of his early converts. Hopfield's work as a recruiter was perhaps the single most important contribution to the early growth of the revitalized field [i.e. neural computing]."

I asked Hopfield about this when I iterviewed him at the California Institute of Technology:

"I don't know if it had anything to do with me personally, or only with the fact that you could now do some mathematics. It would be hard to separate those two. In a certain sense, I had the advantage of being a decently known physicist. I had done solid physics, and I had shown also that I could find things in biophysical molecules than people hadn't found before in an interesting fashion. So I had a fairly good record of having identified problems and given reasonable solutions to them. As a result physicists would take a paper seriously which they would not necessarily have read if written by somebody else." (Hopfield, interview)

The quotations above suggest that Hopfield carried some of his personal prestige as a physicist to the field of neural networks.

---

[71] The comparison between the visibility of Hopfield's contribution and that of other, similar contributions made at around the same time gives some support to this 'credibility hypothesis.' Hinton and Anderson's (1989b, pp. 2-3) pointed out that: "For example, in 1981 Hummel and Zucker circulated a technical report explaining that symmetrically connected networks were an important special case, and that their behavior was governed by an energy function. This report eventually appeared as Hummel and Zucker (1983)."

This type of phenomenon was studied in classical studies in the sociology of science. Warren Hagstrom (1965, pp. 67-68 and 175), for example, indicated that:

> "Distinguished scientists may be able to change specialties and carry their prestige with them. Similarly, scientists may be able to move from a discipline of high prestige to one with less prestige and carry some of their original prestige with them . . . The scientist who is established in a discipline carries some of his prestige with him when he transfers to another."

This type of 'credibility hypothesis' is sometimes linked to other 'sociology of specialties' ideas, such as the belief that the 'dynamics' of emergence of new specialties or 'research networks' explains scientific change, as if scientific specialties had a predetermined cycle of life (e.g. emergence, growth, and decline). Some aspects of the 'branching model' of scientific change developed by Michael Mulkay and colleagues (Mulkay, 1975; Mulkay et al., 1975) point in this 'cycle of life' direction.

> "There seems indeed to be an upper limit between one and two hundred members, beyond which research networks tend to break up into smaller groupings . . . Research networks undergo a continuous process of growth, decline, and dissolution . . . In science new problem areas are regularly created and associated social networks formed. The onset of growth in a new area typically follows the perception, but scientists already at work in one or more existing areas, of unresolved problems, unexpected observations or unusual technical advances, the pursuit of which lies outside their present field. Thus the exploration of a new area is usually set in motion by a process of scientific migration." (Mulkay, 1975, pp. 519-520)

This emphasis on migration could be applied to the case of Hopfield and the arrival of physicists to neural networks that I mentioned earlier, but it would miss important points of the evolution of neural network research in the 1980s. John Law and Barry Barnes (1976) criticised Mulkay's model of branching for not being able to distinguish clearly between the *process* of

innovation and the *consequences* of innovation. In their view, Mulkay's account is useful to study the social consequences of scientific innovation (i.e. the emergence, institutionalisation, and reorganisation of areas of research), but creative activity in science is better accounted for in terms of accepted problem-solutions or exemplars being used as resources for solving new puzzles or interpreting new situations.

Elaborating on Thomas Kuhn's (1970) notion of paradigm (exemplar) and Mary Hesse's (1966) work on the role of models and analogies in science, Barry Barnes (1974, 1982, 1983) developed a scheme for the study of the scientific change (and for the study of the generation and validation of knowledge in general). Barnes (1974, pp. 57 and 87) emphasised the importance of metaphors in scientific innovation:

> "Science may be regarded as a loosely associated set of communities, each using characteristic procedures and techniques to further the metaphorical redescription of a puzzling area of experience in terms of a characteristic, accepted set of cultural resources . . . Creative, non-routine scientific activity becomes intelligible as an aspect of the universal human propensity to create and extend metaphors — a propensity so basic that without it not only would the existence of real cultural change be impossible but also the existence of culture itself. One important way in which creative activity occurs is by a problem or puzzle being seen as an example of an existing exemplar or model problem solution. Kuhn (1970, Postscript) has stressed the key role of this kind of event."

This scheme of metaphorical redescription can be used to analyse the development of Hopfield's (1982) innovation. Later in the section I will use the same scheme to study the development of Ackley, Hinton, and Sejnowski's Boltzmann machine.

John Hopfield is a theoretical physicist at the California Institute of Technology (CalTech) who worked in molecular biology in the 1970s. Towards the end of the 1970s, he was invited by Frank Schmidt to give a talk in one of the Neural

Sciences Research Programme meetings at MIT. Motivated by the lack of theory in neural biology (Hopfield, interview), Hopfield continued to attend those meetings for some years. I asked Hopfield about the origins of his interest in neural networks.

> "I gave a talk on biological molecules. It was obvious listening to that meeting that there were some marvellous problems, that there was a desperate need for theory . . . Relatively speaking, there were no theorists in neural biology." (Hopfield, interview)

Surely researchers who worked in neuroscience-oriented neural networks throughout the 1970s, such as David Willshaw and Christoph von der Malsburg (von der Malsburg & Willshaw, 1977; Willshaw & von der Malsburg 1979) would not agree with Hopfield's last statement in the quotation above.[72] But however Hopfield got interested in neural networks, what interests me here is how he developed his model — and here is where Barnes' work on metaphors helps. Hopfield used a metaphor from statistical physics — the so-called Ising model of magnetic material or 'spin-glass' — in order to develop his 1982 network. Hopfield used the spin-glass as a metaphor for redescribing a class of neural network systems (content-addressable associative memories).

> "In physical systems made from a large number of simple elements, interactions among large numbers of elementary components yield collective phenomena such as the stable magnetic orientations and domains in a magnetic system . . . Any physical system whose dynamics in phase space is dominated by a substantial number of locally stable states to which it is attracted can therefore be regarded as a

---

[72] Paul Smolensky, an important member of the PDP group, suggested some reasons why physicists are sometimes motivated to do research on neural networks (his description could perhaps be applied to the case of John Hopfield):"Physicists are always on the lookout for new places to apply their tools, and they are perpetually attracted to the idea that some very simple model will explain a lot of phenomena once they have formulated and used their mathematical toolket on it. They have a predisposition to believe that the brain can be understood in some deep ways with some simple models that they can understand, and analyse, and prove results about, and so on. The appeal to that community has to do with the idea that this [neural networks] is somehow a theory about the brain, a new place for physics technology to be applied" (Smolensky, interview).

> content-addressable memory. The physical system will be a potentially useful memory if, in addition, any prescribed set of states can readily be made the stable states of the system." (Hopfield, 1982, p. 460)

Associative content-addressable memories have been a subarea of neural network research for a long time. Their origins go back to the late 1950s and early 1960s (Taylor, 1956; Steinbuch, 1961). In the 1970s, neural network researchers like J. A. Anderson, T. Kohonen, and D. Willshaw developed several schemes for associative content-addressable memories.[73] Work on associative memory networks has continued over the years.[74]

John Hopfield reformulated the problem of associative content-addressable memory using Ising ferromagnets as a metaphorical resource. Figure 4-1 (after Hertz et al., 1991, p. 25) represents a simple model of magnetic material.

---

[73] An early example is (Willshaw et al., 1969). Other examples are (Willshaw, 1971), (Kohonen, 1977) and (Anderson, 1972).

[74] (Kanerva, 1988) is a recent development. Beale & Jackson (1990, ch. 8) reviewed associative content-addressable memory (neural network) research.

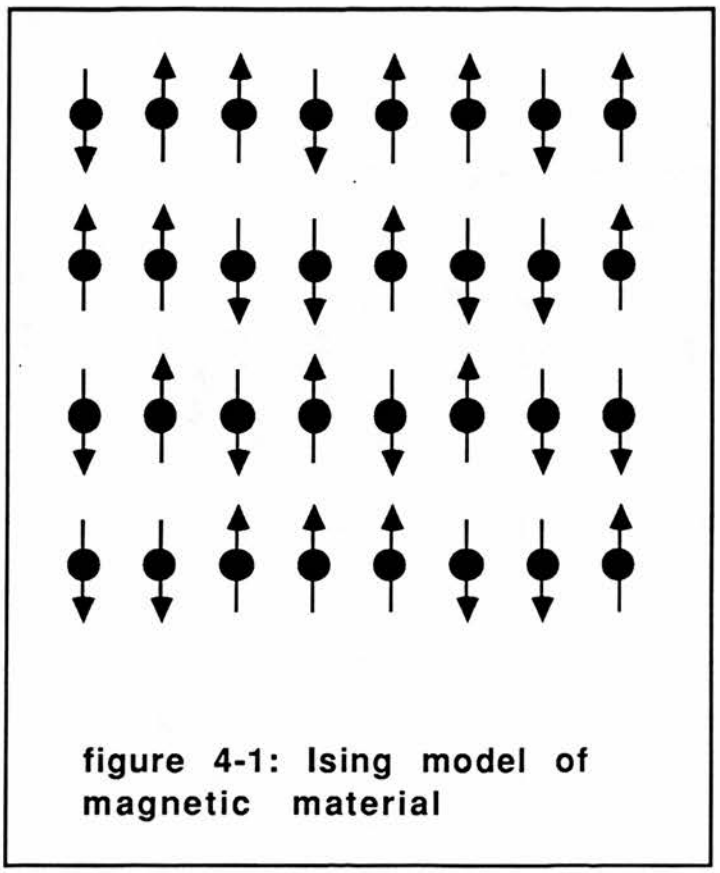


figure 4-1: Ising model of magnetic material

Figure 4-1 shows a set of atomic magnets (spins) arranged on a regular lattice. In this case the spins can only point in two directions (up and down). The dynamics of the magnetic material (called 'spin glass') of figure 4-1 is described as follows (Hertz et al., 1991, pp. 25-26). Each of the spins (S_i) is influenced by the magnetic field at its location (h_i). This magnetic field consists of an internal field produced by the other spins ($\sum_j w_{ij}S_j$) plus any external field (h^{ext}) applied by the experimenter. The magnetic field influencing spin S_i is defined as:

$$h_i = \sum_j w_{ij}S_j + h^{ext}$$

Each spin (S_i) tends to line up parallel to the local field acting on it. Hopfield thought of the spins as 'neurons' (processing units in a neural network), and of the interactions between them as connections in a neural network system. The effect of the external field can be compared with the effect of a threshold.

Thus Hopfield redescribed associative memory neural networks using the spin-glass as a metaphor.

One important characteristic of a magnetic material like the one represented in figure 4-1 is that the exchange interaction strengths between spins are symmetric (that is, $w_{ij}=w_{ji}$). Similarly, Hopfield used symmetric connections in his neural network. One very important property of the spin glass is that its dynamics is defined by an 'energy function.' Assuming that the values of the spins are +1 and -1, and that the external field is the same for all the spins, the (total) energy of a spin-glass over all the spins is:

$$H = -\frac{1}{2}\sum_{i\,j} w_{ij}S_iS_j - h^{ext}\sum_i S_i$$

With $h^{ext}=0$, this is parallel to the energy function suggested by Hopfield (1982, p. 462).

Hopfield (1982) used the spin glass model as a metaphorical resource for developing his neural network. He extended (in Kuhn's and Barnes' sense of the term) the spin glass exemplar (or accepted problem-solution) to a new situation: associative neural networks. In other words, he redescribed associative neural networks in terms of the spin glass model, and in doing so he produced an important innovation.

> "I realised that if you make the system feedback in a symmetric fashion, then its mathematics was controlled, and its mathematics turned out to be the mathematics of the spin glass. So I could bring some things from statistical physics in; otherwise there is little ability to control the mathematics. Being able to get a grasp on how to control the mathematics was really, as far as I was concerned, the big advance that the (1982) paper represented. You can always try to make plausible, or show by simulation, that certain things are possible or might happen, but when you can actually ensure that they will happen, and show why they happened, it produces a serious change in the way that people will view the subject." (Hopfield, interview)

Let me now discuss some details of Hopfield's (1982) neural network. For convenience, I will use binary (1, 0) activation values instead of the (+1, -1) used in the definition of the spin-glass above, and o_i (the output of a unit) instead of the s_i of the spin-glass above. In a Hopfield network every processing unit (or 'neuron') is connected to every other unit except to itself. Figure 4-2 below is a possible graphical representation of a Hopfield network with four units (after Beale & Jackson, 1990, p. 134).

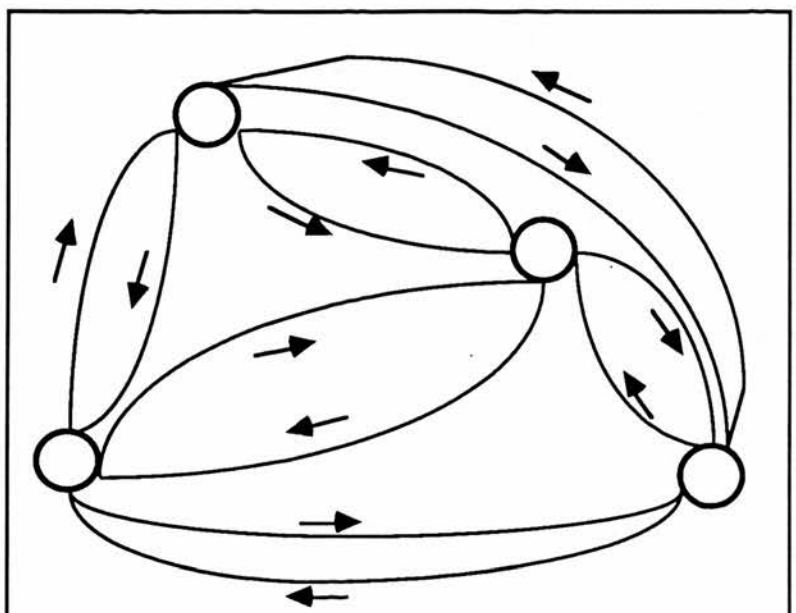


figure 4-2: Hopfield network

In a Hopfield network, the connections between any two units (say unit i and unit j) are symmetric, that is they have the same value (weight) in both directions (from j to i, and from i to j, $w_{ij}=w_{ji}$). The activation function of a unit in a Hopfield network is the usual one in neural network research. A unit (say unit i) fires (sends output activation to other units, i.e. $o_i=1$) if the sum of weighted inputs which it receives equals or exceeds its threshold value (θ_i). If the sum is smaller than its threshold value, then unit i does not fire ($o_i=0$). In symbols:

$$o_i=1 \qquad \text{if } \sum_{j\neq i} w_{ij}\, o_j \geq \theta_i$$

$$o_i=0 \qquad \text{if } \sum_{j\neq i} w_{ij}\, o_j < \theta_i$$

Note the parallelism between this activation rule and the equation that defines the internal field at a given location of the spin glass. The operation of a Hopfield network is asynchronous; the neurons fire in a random order, one at a time. Each neuron makes approximately the same number of attempts to fire per second. So at any instant of time each neuron has roughly the same probability of firing, and over a period of time every neuron will have fired on average the same number of times (Aleksander & Morton, 1990, p. 95). This is also similar to the spin glass situation. At low temperature, each spin tends to line up parallel to the local field acting on it asynchronously in random order (Hertz et al., 1991, p. 26).

As figure 4-2 above shows, unlike the perceptron or other neural network systems, a Hopfield network does not have input units and output units. The units are not structured in layers. An input pattern is fed into the network by setting the initial values of all the units. Then the system is left alone, until it converges into a stable state. This state is the output of the network for that input.

The crucial aspect of Hopfield's (1982) contribution — a consequence of his use of the spin glass metaphor — was the notion of 'energy' of a (symmetrically connected) neural network. The energy of a Hopfield system (a global measure of its performance) decreases every time a unit updates its state (a local operation), until a local minimum (a stable state of the system) is reached. Thus the *local* activity of each unit contributes to the minimisation of a *global* or 'collective' (as Hopfield put it) property of the whole system. Patterns are

stored in local minima of the energy function. One of the most important properties of this type of network is that it can work as a content-addressable memory so that, under the right circumstances, the network will retrieve correct whole patterns when presented with degraded versions of (input) patterns.

The (decreasing) change of energy (ΔE) provoked by unit i when it updates its state can be defined as follows (Aleksander & Morton, 1990, pp. 96-98):

$$\Delta E = - \Delta o_i \left(\sum_{j=1}^{n} w_{ij} o_j - \theta_i \right)$$

where $\sum_{j=1}^{n} w_{ij} o_j - \theta_i$ is the summation carried out by unit i

Every time a processing unit changes its state (this is represented in the equation above as Δo_i) the energy (E) decreases.(i.e. the energy change, ΔE, is negative.[75] The total energy of the system at time t is related to the sum of the energy of all the units, and was defined by Hopfield as follows:[76]

$$E = - \frac{1}{2} \sum_{i\,j} w_{ij} o_j o_i + \sum_{i} o_i \theta_i$$

This energy function corresponds to the equation defining the energy of a spin glass system. It was said earlier that the energy of the spin-glass has a very important property: it has many stable states. Hopfield's idea was to store patterns in those stable states (local minima). Thus stored memory patterns would be attractors of the energy landscape of a Hopfield network. The

---

[75] Aleksander and Morton (1990, p. 96) put this as follows: "Changes in o_i can occur only if o_i is 0 and the so-called activation ($\Sigma w_{ij} o_j - \theta_i$) is positive and Δo_i is also positive, *or* if o_i is 1, in which case the activation must be negative, as is Δo_i. Thus the product $\Delta o_i(\Sigma w_{ij} o_j - \theta_i)$ is always positive . . . which ensures that ΔE is always negative."

[76] The energy of unit i which leads to the energy change ΔE is defined as follows (Aleksander & Morton, 1990, p. 96): $E_i = - o_i \left(\sum_{j} w_{ij} o_j - \theta_i \right) = - \sum w_{ij} o_j o_i + o_i \theta_i$

dynamics of the system can be represented as a ball rolling towards a minimum (a 'valley') of that landscape.

> "Thus, the algorithm for altering V_i [here o_i] causes E [the energy] to be a monotonically decreasing function. State changes will continue until a least (local) E is reached. This case is isomorphic with an Ising model . . . When T_{ij} [here w_{ij}] is symmetric but has a random character (the spin glass) there are known to be many (locally) stable states . . ." (Hopfield 1982, p. 462)

Working with a Hopfield network involves two phases, namely the storage phase and the recall phase. In the storage phase connection weights are assigned so that the patterns that one wants to store in the system are represented by stable states (local minima) of the energy landscape.[77] In the recall phase the system is given an input, and it cycles through a succession of states until it converges to a stable state or local minimum of energy (this state is the output for the given input). Thus a Hopfield network can work as a associative content-addressable memory. Hopfield (1982, p. 462) indicated that his network can store approximately 0.15N patterns, where N is the total number of units ('neurons') of the network.[78]

Hopfield's (1982) network has some important limitations. Hinton and Anderson (1989b, pp. 3-4) described them in the following terms:

> "Unfortunately, there is no guarantee that it [the Hopfield network] will settle to the *nearest* energy minimum, and

---

[77] For details about the storage and the recall phase see Beale and Jackson (1990, ch. 6) and Hertz et al. (1991, ch. 2). For a system with bipolar (+1, -1) values and with thresholds of the units being zero, connection weights are assigned as follows (Beale & Jackson, 1990, p. 136):

$$w_{ij} = \sum_{s=0}^{M-1} x_i^s x_j^s \quad \text{if } i \neq j$$

$$w_{ij} = 0 \quad \text{if } i=j,\ 0 \leq i,\ j \leq M-1$$

w_{ij} is the value of the weight of the connection between unit i and unit j. x_i^s is the i-th element of the exemplar pattern for class s. There are M patterns, from 0 to M-1.

[78] This storage capacity is today seen as quite small, and the Hopfield network is being developed in various directions. For these developments, and related work in optimisation problems see Hertz et al. (1991, ch. 3-4).

only a very few vectors can be stored without creating spurious local minima."

Nevertheless, Hopfield's network was developed in several directions (see previous footnote), and the study of Hopfield nets became a subarea of research within neural computing. Hopfield (1984) himself carried out one of the earliest modifications of his model. He substituted a continuous activation function (a nonlinear, sigmoid activation function) for the step function of his original 1982 system, and claimed that the properties of this new type of network were still very similar those of the 1982 model.[79]

Hopfield was a consultant at AT&T Bell Laboratories (Murray Hill, New Jersey), where Larry Jackel and Richard Howard built an analog microchip of Hopfield's (1984) model (Larson 1986, pp.114-116). It was a twenty two-neuron (484 weight) chip, probably the first neural network chip ever built. Hopfield described it as follows:

> "The chip was never used for anything. It wasn't big enough. It only had 22 neurons on it, and it was sort of a *tour de force*. But it was nice to see that you can translate [mathematics into a chip]. Mathematics put into a chip is not literally exactly like that [like 'pure' mathematics], it is only sort of figuratively like that. The important thing is that the real physical system actually behaves the way that you would expect it to do." (Hopfield, interview)

Larson (1986, p. 116) reported on the views of chip-builders Jackel and Howard:

> "Their first chip, built this year [1986], had 22 'neurons' and 484 'synapses.' It stored four short names and acted as a small associative memory . . . The system worked exactly as Hopfield's simulations had predicted. 'The big difference

---

[79] Sigmoid functions had been used earlier by neural network researchers (e.g. Grossberg, 1976b). Sigmoid functions are s-shaped,continuous nonlinear functions. The main characteristic of the sigmoid function can be described as follows (Anderson & Rosenfeld, 1988, p. 244). At low and high levels of activation, changes of activation lead to very small changes in output. At intermediate range levels of activation, changes in activation lead to (comparatively) large changes in output.

between this and a simulation is that a simulation takes about one second,' says Jackel. 'This takes a *millionth* of a second.' They are now testing their latest chip. It has 512 neurons and half a million interconnections."

Several neural network chip implementations of Hopfield models have been (and are been) built at Bells Labs and elsewhere since then, using both analog VLSI and optical hardware. Hardware implementation is today an important area of neural computing research, and developments after Hopfield's models were among the first steps in that direction. It is interesting to note that Hopfield insisted on the importance of hardware implementation in his 1982 paper.

> "The model could be readily implemented by integrated circuit hardware. The conclusions suggest the design of a delocalized content-addressable memory or categorizer using extensive asynchronous parallel processing . . . Implementation of a . .. . model by using integrated circuits would lead to chips which are much less sensitive to element failure and soft-failure than are normal circuits . . . Their asynchronous parallel processing capability would provide rapid solutions to some special classes of computational problems." (Hopfield, 1982, pp. 460 and 464)

Some early applications of the Hopfield network created considerable excitement. An example is Hopfield and Tank's (1986) application of the Hopfield network to the 'travelling salesman' problem (a NP-complete problem, i.e. the steps required to solve it grow exponentially with the size of the problem). Although this application received much more critical evaluations later (see Darpa, 1988, pp. 197-198), the Hopfield network is being developed and applied in a number of directions (as I said earlier, the study of Hopfield networks has become a subarea of research in neural networks).[80]

---

[80] The travelling salesman problem is an optimisation problem where, given a list of cities and the distances between each pair of cities, one wants to find the minimum-length tour that visits each city exactly once. Durbin and Willshaw (1987) claimed that they got better results using a different neural network algorithm called the 'elastic net' (based on Willshaw and von der Malsburg's earlier work on topographically ordered mapping of connections in the brain).

But the importance of Hopfield's contribution in the emergence of neural computing in the 1980s does not stop there. David Ackley, Geoffrey Hinton, and Terrence Sejnowski developed the Hopfield network further and combined it with other elements in their 'Boltzmann machine' network (Hinton, Sejnowski, & Ackley, 1984; Ackley, Hinton, & Sejnowski, 1985; Hinton & Sejnowski, 1986).

In the rest of this section I examine the development of the Boltzmann machine network. I will show that the 'metaphor scheme' (Barnes, 1974) is useful for the study of the development of the Boltzmann machine innovation too. The Boltzmann machine (BM) network was very important for the history of neural networks, because for the first time a solution was given to the problem of training multilayer systems. The neural network controversy did not reopen completely until a later solution to that problem was given (namely back-propagation technique, studied in section 5.2). However, the BM system was a very important step towards that reopening, and it encouraged researchers to develop techniques for training multilayer systems with different architectures.

The origins of the BM network go back to the meetings of researchers from the PDP group and elsewhere in the early 1980s (see section 4.1). John Hopfield was invited to give a talk in one of these meetings, and it seems that this personal contact helped Hinton and Sejnowski conceive the idea of their BM network. Hopfield told me about that:

> "There was a meeting at the University of Rochester in the summer of 1982 or so. I went and talked. Hinton and Sejnowski were there, saw the energy function, went off, and the Boltzmann machine paper was done very shortly after that." (Hopfield, interview)

Jerome Feldman, the organiser of that meeting, recalls that moment too. It is interesting to point out that at that point some

of the participants in the meeting did not know about Hopfield's (1982) work. Here is how Feldman remembers the episode:

> "Like in the second meeting, 1982 or so, we had heard about what Carver Mead was doing, invited him, and he couldn't come, but this guy we had never heard of called John Hopfield came, and he talked about stuff that seemed very strange, but that's when Sejnowski and Hinton got very excited about Boltzmann machines. It was in the middle of that meeting that they had this very exciting idea, and they ran off. It was all very interesting and very exciting." (Feldman, interview)

Hinton and Sejnowski had studied computational problems in vision (Ballard, Hinton, & Sejnowski, 1983), and thought that a neural network-like 'relaxation scheme' of parallel computation (i.e. computation as a process of satisfaction of a large number of weak constraints) could be interesting for those problems.

> "A visual system must be able to solve large constraint-satisfaction problems rapidly in order to interpret a two-dimensional intensity image in terms of the depths and orientations of the three-dimensional surfaces in the world that gave rise to that image." (Hinton & Sejnowski, 1986, p. 282)

Hopfield's (1982) system was an interesting scheme for this type of 'relaxation search.' But one limitation of this system is that patterns are stored in local minima of the energy landscape. Ackley, Hinton, and Sejnowski developed further Hopfield's notion of energy so that *global* energy minima (more optimal solutions) could be used to store patterns.

Hinton and his colleagues introduced a new parameter in their Hopfield-like network, namely the so-called 'temperature' of the system. The notion of temperature goes back to the spin-glass analogy used by Hopfield. If the temperature of a spin glass is not very low, thermal fluctuations tend to change the state of the spins, and thus to upset the tendency of each spin to align with its field (Hertz et al., 1991, p. 26). These fluctuations decrease as temperature is lowered. So by raising the

'temperature' of a Hopfield-like system, it could jump out of a local minimum (to a state of higher energy), and then a better minimum could be searched. If one thinks of the state of the system as a small ball in the energy landscape, the way to get the ball out of a local minimum would be to 'shake' it a little (to rise the 'temperature').

Hinton and Sejnowski introduced the idea of 'temperature' in their system, and used stochastic units instead of the deterministic ones used by Hopfield (and many other neural network researchers before). 'Temperature' in a BM system is not physical temperature, but level of noise (so to speak). The higher the temperature, the greater the likelihood that the system will jump to a state of higher energy (thus violating the behaviour typical of a Hopfield net, which always goes to a state of lower energy). The stochastic activation function defines the probability that a unit in a BM network will fire (i.e. will produce output 1) as follows:

$$p_k = \frac{1}{1+e^{-\Delta E_k/T}}$$

Figure 4-3 below shows the effect of the temperature parameter (T) on the sigmoid activation function (p_k) used in the BM.

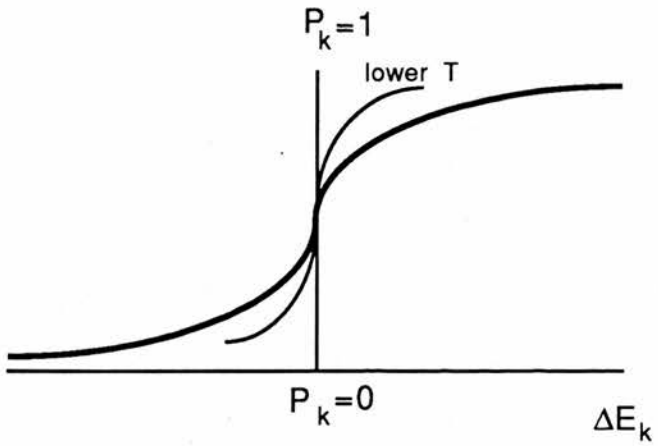


figure 4-3: sigmoid function

As the sigmoid function (in bold) shows, the greater the activation of a unit (ΔE_k, see below), the greater the probability that this unit will fire (p_k). As the temperature is lowered (see figure 4-3), the activation function approaches the deterministic, step activation function (the one originally used by Hopfield in his 1982 paper). If temperature is increased, the sigmoid function 'flattens' (so to speak), and it becomes easier for a unit to violate the deterministic rule. This allows the system to go to states of higher energy (and jump from local minima).

Like in a Hopfield network, the units of a BM network have binary activation values (1, 0), update their state asynchronously in random order one at a time, and the connections between them are symmetric. The energy function of the BM network is the same as that of a Hopfield network (and therefore corresponds to the energy of a spin-glass). In a BM network each time a unit updates its state, it minimises the global energy of the system. That is:

$$\Delta E_k = \sum_i w_{ki} o_i - \theta_k$$

Ackley, Hinton, and Sejnowski called their system the 'Boltzmann machine' because, if the activation rule p_k defined above is used, the probability of finding the system in a particular state is defined by the Boltzmann-Gibbs distribution from statistical physics.[81]

> "The decision rule [p_k, see above] is the same as that for a particle which has two energy states. A system of such particles in contact with a heat bath at a given temperature will eventually reach thermal equilibrium and the probability of finding the system in any global state will then obey a Boltzmann distribution." (Ackley, Hinton, & Sejnowski, 1985, p.640)

The Boltzmann-Gibbs distribution is defined as follows (Hertz et al., 1991, pp. 275-277). If a physical system has a set of states α, and each of the states has an energy H_α, then at thermal equilibrium each of the possible states α occurs with probability:

$$P_\alpha = \frac{1}{Z} e^{-H\alpha/k_B T}$$

'T' is the temperature of the system, k_B is Boltzmann's constant, and 'Z' is the normalising factor:

$$Z = \sum_\alpha e^{-H\alpha/k_B T}$$

Because the 'temperature' in a BM is not related to physical temperature, k_B can be taken to be 1 (and therefore can be eliminated from the Boltzmann distribution above). Hinton and

---

[81] Austrian physicist Ludwig Boltzmann (1844-1906) made important contributions to the atomic theory of gases and to statistical mechanics. He also introduced the basic relation connecting entropy with the number of accessible states of a physical system. Josiah W. Gibbs (1839-1903) was a physicist from the United States who made important contributions to thermodynamics and statistical mechanics.

his colleagues used the properties of the Boltzmann distribution to derive their learning algorithm.

In a BM, in order to find global minima, the temperature of the system is decreased using a technique called 'simulated annealing.'[82] This method consists of starting at a high temperature and reducing it slowly until the system settles into a solution. At high temperatures, jumps from lower to higher energy states are allowed, and thus local minima states can be escaped. As temperature is reduced, the probability of changing from a lower to a higher energy state decreases.

One interesting property of the Boltzmann distribution is that, at thermal equilibrium (when the probabilities of the states no longer change) the relative probability of two global states is determined only by their energy difference, so that lower energy states are more probable than higher energy ones.[83] Hinton and his colleagues took advantage of the properties of the Boltzmann distribution to develop their BM learning algorithm.

A very important aspect of Hinton and colleagues' BM network is that they defined its architecture in a way which allowed them to solve the problem of training hidden units in a multilayer neural network. They divided the units of the network into three classes: input units, hidden units, and output units, as figure 4-4 below (after Beale & Jackson, 1990, p. 151) shows.

---

82 The 'simulated annealing' technique was introduced by Kirkpatrick, Gelatt, & Vechi (1983) for solving optimisation problems with conventional computers. Geman & Geman (1984) studied the annealing schedule (Hinton & Sejnowski, 1986, pp. 287-289).

83 The relative probability of two global minima states at thermal equilibrium (a consequence of the Boltzmann distribution) is:

$$\frac{P_\alpha}{P_\beta} = e^{-(E_\alpha - E_\beta)/T}$$

If P_α is the probability of being in the α-th global state, and P_β is the probability of being in the β-th global state, and if E_α is a lower energy state than E_β, then $P_\alpha > P_\beta$ (at thermal equilibrium) (Hinton & Sejnowski 1986, p. 289). "The Boltzmann distribution has some beautiful mathematical properties . . . In particular, the difference in the log probabilities of two global states is just their energy difference (at a temperature of 1) [at this temperature the sigmoid function has the shape in bold in figure 2-14]" (Ackley, Hinton, & Sejnowski, 1985, p. 640).

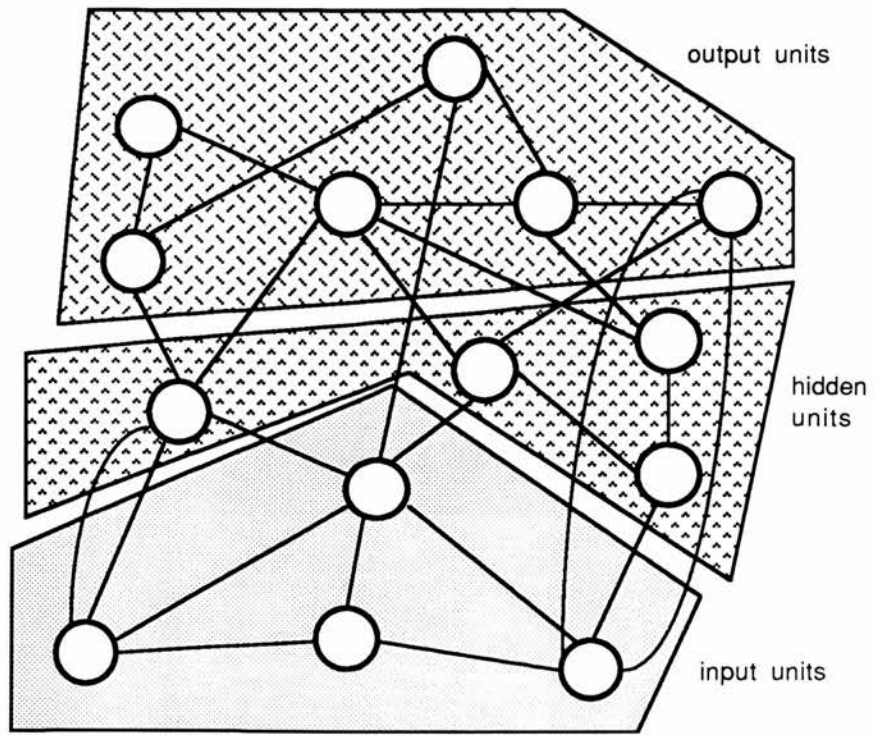


figure 4-4: Boltzmann machine network

In figure 4-4 many of the connections of the BM network are not shown. The connections between the units are Hopfield-like (symmetric, and in principle every unit is connected to every other unit except to itself). Therefore, the BM network is not divided into layers in the same sense as multilayer feedforward networks are (e.g. in the BM of figure 4-4 there are connections between input units and output units).

Hinton and Sejnowski developed their Hopfield-like with stochastic units before they developed their BM learning technique[84] The relationship between the probability of a global state and its energy (a property of the Boltzmann distribution)

[84] In Hinton & Sejnowski (1983), as Hertz et al. (1991, p. 32) pointed out. This coincides with Feldman (interview) and Hopfield's (interview) accounts (see above) about the meeting in which Hinton and Sejnoswki first conceived of the BM idea. They developed the learning algorithm later, and it seems that they were amazed that the learning algorithm could be developed so 'easily.'

and the relationship between the energy and the weights (as defined by Hopfield's energy function) allowed Hinton and colleagues to develop their technique for the modification of the weights of the BM network.

The BM learning algorithm has two phases. In the first (or incremental) phase, the input and output units are clamped to their correct values (the input/output vectors that the experimenter wants to associate). The system then cycles through its states (the temperature being decreased gradually) until the hidden units reach thermal equilibrium. Then, the weights that connect two units that are both on (their output is 1) are incremented. In the second phase (decremental), only the input units are clamped. The system cycles through its states as in phase 1, until thermal equilibrium is reached. Then the weights between two units which are both 'on' are decremented. The two phases are repeated until the weights become stable. By using this gradient descent method, the best (deepest) global minima are usually reached (Beale & Jackson, 1990, p. 150).[85] The main problem of the BM learning technique is that it is very slow and computationally intensive. Later the behaviour of a BM has been approximated by using a faster technique from physics called 'mean field theory' and real-valued units (see Hinton & Anderson, 1989b, pp. 4-5; for variations of the BM see also Hertz et al., 1991, pp. 169-172).

Despite its limitations the BM network and learning algorithm were of great importance for neural network research. It was the

---

85 Hinton and Sejnowski (1986) summarised their BM contribution in this way: "We have presented three ideas. (i) Networks of symmetrically connected, binary units can escape from local minima during a relaxation search by using a stochastic decision rule. (ii) The process of reaching thermal equilibrium in a network of stochastic units propagates exactly the information needed to do credit assignment. This makes possible a local learning rule which can modify the weights so as to create new and useful feature detectors. The learning rule only needs to observe how often two units are both active (at thermal equilibrium) in two different phases [incremental and decremental]. It can then change the weight between the units to make the spontaneous behavior of the network in one phase mimic the behavior that is forced on it in the other phase. (iii) The learning rule tends to construct distributed representations which are resistant to minor damage and exhibit rapid relearning after major damage . . ." (p.313).

first solution given by connectionist researchers to Minsky and Papert's (1969, pp. 231–232) famous challenge about the problem of learning in multilayer networks. Hinton and his colleagues presented their contribution as a response to that challenge.

> "This 'credit-assignment' problem was what led to the demise of perceptrons (Minsky & Papert, 1969; Rosenblatt, 1962). The perceptron convergence theorem guarantees that the weights of a single layer of decision units can be trained, but it could not be generalized to networks of such units when the task did not directly specify how to use all the units in the network. This version of the credit-assignment problem can be solved within the Boltzmann machine formulation." (Ackley, Hinton, & Sejnowski, 1985, p. 641)

Sejnowski (interview) made a similar comment about Minsky and Papert's (1969) challenge:

> "In the Boltzmann machine, Geoff [Hinton] and I found a learning algorithm which overcame the conjecture by Minsky and Papert that you couldn't generalise the perceptron learning algorithm to a multi-layered architecture. The Boltzmann machine is a generalisation of the perceptron to more than one layer. It's interesting, it turned out that the key assumptions you had to change were two things. First that there are feedback connections *à la Hopfield*, so you have symmetric connections, so it's not longer feedforward net but symmetric net with feedback connections. And second of all, the perceptron was a deterministic machine, whereas the Boltzmann machine was probabilistic. So you make those two changes, and then suddenly it's a completely different architecture, suddenly you can prove theorems, you can discover learning algorithms, you can solve problems that the perceptron couldn't."

John Hopfield emphasised the importance of the learning aspect of the BM contribution, rather than the mere introduction of the temperature parameter:

> "The idea of 'temperature' was a trivial addition [to my (1982) model], present in early drafts of my (1982) paper

> until space required its removal. The thing which was *very* imaginative on their part was to understand that there is a learning algorithm which would train units which were between your input and your output, the so called hidden units . . . The perceptron didn't go any further than it did because there was no way of training hidden units. With the Boltzmann machine you could actually train the hidden units." (Hopfield, interview)

Thus the BM was the first solution to the 'reverse salient' (Hughes, 1983) of training multilayer networks. The transformation of that reverse salient into a 'critical (solvable) problem' (ibid.) has been studied in this section using a 'metaphor scheme' (Barnes, 1974).

The PDP researchers were very encouraged by the BM network. For them it was a successful breakthrough which showed that some of the most important limitations of the neural network systems of the 1960s could be overcome. The BM network was a very important development for the reopening of the perceptron controversy (and the more general debate between the symbolic and the neural network approaches). However, in my view, the 'complete' reopening of the controversy happened a little later, coinciding with the development of the back-propagation network by PDP researchers D. Rumelhart, G. Hinton, and R. Williams (1986) (see section 5.2).

But before I look at the development of back-propagation I would like to discuss briefly some relationships between the Boltzmann machine and certain ideas from cybernetics. In section 2.1 it was seen that the brain/machine issue was of central importance in cybernetics. Brain and machine were compared in terms of information-processing. The interesting thing about cybernetics is that a variety of approaches to information-processing were studied (some more deeply than others). These included symbolic AI (based on the symbol-processing uses of the von Neumann computer), biological control theory, bionics, neural modelling, 'brain theory' (today's computational neuroscience), and neural networks.

Interestingly, John von Neumann himself studied rather 'non-von Neumann' (so to speak) approaches such as reliable automata with unreliable parts and cellular automata. Von Neumann's concern with malfunction in information-processing systems — and therefore with a model of information-processing different from the one based on formal logic — was taken seriously by Frank Rosenblatt in the design of the perceptron. In his original paper on the perceptron Rosenblatt (1958a, pp. 93-94) emphasised the importance of reliable computation with imperfect components (this was one of the reasons for the random connections in his machine):

> "A relatively small number of theorists, like Ashby (1952) and von Neumann (1951, 1956), have been concerned with the problems of how an imperfect neural network, containing many random connections, can be made to perform reliably those functions which may be represented by idealized wiring diagrams. Unfortunately, the language of symbolic logic and Boolean algebra is less well suited for such investigations. The need for a suitable language for the mathematical analysis of events in systems where only the gross organization can be characterized, and the precise structure is unknown, has lead the author to formulate the current model [the perceptron] in terms of probability theory rather than symbolic logic."[86]

It is interesting to note that von Neumann made explicit comparisons between information processing and Boltzmann's statistical physics. One interest of von Neumann was information-processing in systems reliable in case of error or malfunction of some of their components. Von Neumann pointed out that for this an approach different from formal logic-based computation was required. In the 1948 Hixon symposium on 'Cerebral mechanisms in behaviour' von Neumann said (Jeffress, 1951, p. 17):

---

[86] Rosenblatt (1958a, p. 93) was against "brain models which amount simply to logical contrivances for performing particular algorithms (representing 'recall,' stimulus comparison, transformation, and various kinds of analysis) in response to sequences of stimuli — e.g. . . . McCulloch & Pitts (1943) . . . Minsky (1956)."

> "The operations of logic . . . will all have to be treated by procedures which allow exceptions (malfunctions) with low but non-zero probabilities. All of this will lead to theories which are much less rigidly of an all-or-none nature than past and present formal logic . . . In fact, there are numerous indications to make us believe that this new system of formal logic will move closer to another discipline which has been little linked in the past with logic. This is thermodynamics, primarily in the form received from Boltzmann, and is that part of theoretical physics which comes nearest in some of its aspects to manipulating and measuring information."[87]

William Aspray (1990, p. 319) indicated that in his 1952 California Institute of Technology lectures von Neumann was even more explicit in his comments on the relationship between information processing and statistical physics, and related Claude Shannon and Leo Szilard's definition of information to the concept of entropy:

> "An important aspect about this definition [$I = \sum_{i=1}^{n} p_i \log_2 p_i$, see below] is that it bears close resemblance to the statistical definition of the entropy of a thermodynamical system. If the possible events are just the known possible states of the system with their corresponding probabilities, then the two definitions are identical. Pursuing this, one can construct a mathematical theory of the communication of information patterned after statistical mechanics . . . The closeness . . . between information and entropy is inherent in L. Boltzmann's classical definition of entropy . . ." (von Neumann, in Aspray, 1990, p. 319)

Assuming a physical system with a set of states α, entropy (S) can be defined as the width of the probability distribution P_α;

---

87 In 1948 Norbert Wiener talked about statistical mechanics in his definition of cybernetics (in his 1948 'Cybernetics' book). Wiener said that the new science of cybernetics was based on "the essential unity of the set of problems centering about communication, control, and *statistical mechanics*, whether in the machine or in living tissue" (quoted by Aspray, 1990, p. 209, emphasis added).

that is, the more states α that have an appreciable probability, the larger S:

$$S=-\sum_{\alpha} P_{\alpha} \log P_{\alpha}$$

Entropy is interpreted in information theory as follows (Hertz et al., 1991, pp. 277-279):

$$I=-\sum_{\alpha} P_{\alpha} \log_2 P_{\alpha}$$

Information entropy (I) is written with a base 2 logarithm in order to give the result in bits, and is the average amount of information required to specify one of the states of the system.[88]

Thus the origins of the differences between the statistical thermodynamics-like approach to information-processing of the Boltzmann machine and the formal logic-based approach which is at the basis of symbolic AI go back to the time of cybernetics. Interestingly, McCulloch and Pitts (1943) formal neurons were used in contributions related to both approaches. Marvin Minsky (1967), who used McCulloch-Pitts neural nets as a 'language' in his study of the theory of computation, made the following comments about the relationship between statistical thermodynamics and information processing:

---

88 The reason for the 'minus' sign in the equation above is that the logarithm of a number less than one is negative. When all the states (K) are equally likely entropy information is defined as $I=\log_2 K$, which is the number of bits needed to specify one choice out of K alternatives. For example, in the case of 8 alternatives, it is easily shown that 3 bits are needed to specify one choice ($\log_2 8=3$). Let us say that the choice is 100. Then, in the table below, the first bit (1) eliminates choices (v) to (viii). The second bit (0) eliminates choices (i) and (ii). And the third bit (0) eliminates choice (iii), and therefore the desired choice (iv) is obtained.

(i) 111
(ii) 110
(iii) 101
(iv) 100
(v) 011
(vi) 010
(vii) 001
(viii) 000

> "This [theory of computation] really is a branch of mathematics that stems directly from non-numerical, logical foundations . . . In developing a Theory of Computation, we are trying to deal with systems composed of a great many parts, or very intricate structures. Classical mathematical methods can do this only in very special situations, and their limitations are very serious. Classically, one is unable to cope with even a few simultaneous non-linear equations, to say nothing of a few dozen, or a few million. Now it is true that under certain special conditions mathematical analysis can 'revive,' as it were, when the situation gets complex in such a way that the parts of the system can be treated as individually and independently random — this is what happens in Statistical Thermodynamic theories. But it must be stated, explicitly and emphatically, that this is just what does *not* happen when, as in a computation system, the structure has a more organized, purposeful structure. The statistical analysis works beautifully for things like gases. It works for precious little else. There simply is no reason to suppose that, as computations grow large, one will discover anything non-trivial by trying to 'average-out' the effects of many events. The effect of the 'conditional' . . . is too strong to allow anything like a 'conservation' concept to have a place in the theory. Fortunately, the systems of computation have other features that make possible some analysis, though of a very different kind. Instead of the *statistically defined* events used in physics, we use *logically-defined classes* of computations or expressions." (Minsky, 1967, p. x)

Minsky's view represents the dominant approach to information-processing in AI research since the perceptron debate was brought to closure. But the re-emergence of neural network research in the second half of the 1980s reopened the old brain/machine question of cybernetics (with its variety of approaches). Hinton and Sejnowski's BM neural network system is especially interesting in this context. The BM can be seen as a re-emergence of some of the information-processing ideas of cybernetics, and in particular of the ideas about the relationship between information-processing and statistical physics. Ackley,

Hinton, and Sejnowski (1985) were aware of this. They ended their paper with a quote from von Neumann's talk at the Hixon symposium (see above), and argued "that the Boltzmann Machine is a simple example of a class of interesting stochastic models that exploit the close relationship between Boltzmann distributions and information theory" (p. 647).

In this section I have shown that the development of both Hopfield's network and Hinton and colleagues' Boltzmann machine network can be studied using a 'metaphor scheme.' I have also shown that the Boltzmann machine — a solution to the problem of training multilayer networks — was an important antecedent of the reopening of the neural network controversy. I have insisted on the difference between (i) these innovations and the process of enrolment of allies and resources that finally brought about the reopening of the neural network controversy on the one hand, and (ii) the consequences of those developments (migration of scientists and engineers to neural networks, reorganisation of the field) on the other. The growth and institutionalisation of neural network research will be studied in section 5.3. Before that, in the coming sections (mainly in section 5.2) I look at another innovation of great importance for the reopening of the neural network controversy in the 1980s: the back-propagation learning technique for multilayer, perceptron-like networks.

◆ FIVE

Controversy Reopens

5.1 History of back-propagation

I start this section by discussing the sense in which Minsky and Papert's (1969) critical study of Rosenblatt's perceptron can be said to have been a 'positive contribution' to neural computing. The development of the back-propagation technique can be understood within this context. Afterwards I study some aspects of the history of back-propagation. In particular, I look at Paul Werbos' unsuccessful attempts to apply his 'dynamic feedback algorithm' to neural network research. Surprisingly, Marvin Minsky himself was (once again) a protagonist of some of these historical developments. I conclude by criticising the notion of discovery that is implied in some of Werbos' claims, and by indicating that in the 1970s he was not powerful enough to overcome the resistance that he found to the very idea of applying a back-propagation-like technique to neural networks. The closure of the perceptron controversy had not been revised yet.

In sections 3.3 and 3.4 I showed that Minsky and Papert's (1969) study was a very important element in the closure of the perceptron controversy. Minsky and Papert had criticised neural networks on similar grounds earlier in the 1960s, and that type of criticism had had an effect on the crisis of neural network research in the mid-1960s. Minsky and Papert's (1969) study was the last 'push,' as it were, which brought the perceptron controversy — and the more general debate between the symbol-processing and the neural network approaches to AI — to closure. Minsky and Papert's (1969) study was the result of a costly (in terms of time and research effort) project in which Rosenblatt's perceptron was 're-enacted' (section 3.1). Minsky and Papert dedicated much time and effort to their 'Perceptrons' book ('we got too greedy,' were Minsky's words). As a consequence,

publication was delayed, and by the time the book came out it was quite late in the controversy. Timing is an important rhetorical element, and by publishing late in the controversy Minsky and Papert were successful in getting a 'last word effect' with their book.[89] Pamela McCorduck (1979, p. 88) indicated this 'last word effect' in her history of AI:

> ". . . Papert . . . who was to become Minsky's research partner in a number of efforts, including the *last word* (so far) in the subject [neural networks], a book called 'Perceptrons' (Minsky & Papert, 1969)." (emphasis added)

But, by being published quite late in the controversy, Minsky and Papert's (1969) study had another effect on the evolution of neural network research. Although in the short term the study was the last push for the closure of the perceptron controversy, in the long term the weak points of neural network research became clearer. In this sense, it can be said that Minsky and Papert made a 'positive contribution' to neural network research. Emphasising where your opponents' problems lie is in a sense a positive contribution to the position you are criticising. In addition, by publishing such a detailed and elaborated study about the position they were criticising, Minsky and Papert created some interest in neural networks. Bernard Widrow (interview) confirmed this:

> "What the book [Minsky & Papert, 1969] did when it came out was that it did create quite a bit of interest in neural nets in a negative way. There was more interest as a result of the book than before . . ."

Harry Collins' (1985, p. 150) comment on the effect of strong criticism applies here:

> ". . . The first ploy (and an extremely effective one) of a sensible and determined scientific critic . . . is to ignore the contentious claim. Even to criticize an idea in a

---

[89] For an example of the importance of the 'final word' see Star (1989a, p. 150).

devastating way is to start to bring about its institutionalization."[90]

Minsky and Papert did not ignore their opponents' claims. In section 3.1 it was said that in their 1969 'Perceptrons' study, Minsky and Papert 're-enacted' (Latour, 1987, pp. 60-61) Rosenblatt's perceptron results. In Latour's scheme, Minsky and Papert's (1969) study of Rosenblatt's perceptron belongs to the 'third way' of reading a scientific paper. 'Giving up' is the most usual way of 'reading' a scientific text: just not reading it. The second one, a quite rare one, is 'going along:' the reader believes the author's claim and uses it. The reader refers to it, and by doing so the claim becomes more of a fact. The third way of reading a paper, an extremely rare and costly one, is 're-enacting' everything the author went through, and then — Latour pointed out — at least one flaw is always found even in the best scientific text. Minsky and Papert embarked on a reading of Rosenblatt's work on single-layer neural networks of this third sort, and they not only found one flaw, but many (or so they claimed; see section 3.3). It was very costly for them: it took years of research. But, more importantly, it became even costlier for anybody to challenge their results. This took many more years, almost two decades.

But, by dedicating so much attention and research efforts to criticising the perceptron, Minsky and Papert created some interest in it, and 'constructed' a challenging problem (or reverse salient) for future neural network researchers. This problem could be stated as follows: 'We have showed that single-layer neural network machines have important limitations; the only solution would be to use multilayer systems, but the learning issue in these systems is hopeless.' It was hopeless for Minsky and Papert, but anybody willing to reopen the controversy knew where he or she had to start from. Thus in this sense Minsky and Papert's (1969) criticism could be said to be a positive

---

[90] In this quotation Collins is talking about the early stages of a controversy, but his comments can be applied to Minsky and Papert's both early and late involvement in the perceptron controversy.

contribution to neural network research: it created a reverse salient for future researchers. I showed in earlier sections that early neural network researchers were aware of the problem of training multilayer systems, and had tried to solve it. But they had many other problems. Because of the importance of Minsky and Papert's (1969) study in the closure of the perceptron controversy, the importance of the reverse salient of training multilayer systems became even clearer. In section 3.3 it was seen that Minsky and Papert did not say much about multilayer systems in their study. That is precisely why the only thing they said — their pessimistic intuitive judgement about training those systems — became so important as a challenge for future neural network researchers.

Thus it can be said that neural network research was shaped through the controversies with its critics. This could be related to Susan Star's comments about the socialising aspects of controversy. Elaborating on some ideas of German sociologist Georg Simmel, Star (1989a, p. 121) indicated that:

> "Conflict itself is socializing, in that participation in conflict reflects and develops commitments to certain paths of action. After a long period of time, the very density and durability forms a structure of its own. Battle lines are drawn, and the topics of debate provide a wellspring of problems for research and publication. Career directions are defined with reference to the debate."

This can be related to Harry Collins' remark that 'criticising an idea in a devastating way is to start to bring about its institutionalization' (see above). By emphasising the problem of learning in multilayer systems, Minsky and Papert helped create a reverse salient in neural network research. The 'construction' of a problem is a positive contribution because it reduces the complexity of a situation. Star (1989a, p. 189) underlined the importance of this 'reduction of complexity' in science:

> ". . . Any set of scientific tasks involves multiple problems, qualifications, exigencies, demands, and audiences. To work without getting lost in endless contingencies, scientists

> must draw boundaries and exclude some kinds of artifacts and complications from consideration . . . Part of doing science is transforming problems with many contingencies into those simple enough to work on. Creating well-structured problems requires ignoring complexity . . ."

A good part of the work done in neural networks in the 1980s was directed towards solving the problem of training multilayer systems. In section 4.2 I studied the development of a solution to this problem, namely Ackley, Hinton, and Sejnowski's (1985) Boltzmann machine network. The Boltzmann machine encouraged other researchers to develop techniques for multilayer neural networks with other architectures. In 1986, Rumelhart, Hinton, and Williams, from the PDP group, developed the back-propagation (BP) technique. Soon the BP network became the most successful neural network of the late 1980s. In section 5.2 I will look at Rumelhart, Hinton, and Williams (1986) work on BP, and its importance for the reopening of the controversy. It will be seen that, in the same way as Minsky and Papert's (1969) study was the 'last push' to close the perceptron controversy Rumelhart, Hinton, and Williams' (1986) BP network was the 'last push' to reopen it.

But the idea of back-propagation has more history than what is usually thought. In the rest of this section I look at some earlier (and not so successful) attempts of developing neural networks with back-propagation. I said earlier that, after Minsky and Papert's (1969) study, the importance of the reverse salient of training multilayer networks became even greater than before (as compared to other problems of single-layer networks). It is not surprising, then, that some researchers tried to solve it before Rumelhart and his colleagues carried out their successful work in 1986.

Seymour Papert (1988) criticised Rumelhart, Hinton, and Williams' (1986) work on BP (to be studied in section 5.2), and made an interesting remark. He said that the recent influential work on BP 'could have easily been done twenty years ago:'

> "The influential recent demonstrations of new networks all run on small computers and could have been done in 1970 with ease . . . The examples discussed in the literature are still very small . . . The entire structure of recent connectionist theories might be built on quicksand: it is all based on toy-sized problems with no theoretical analysis to show that performance will be maintained when the models are scaled up to realistic size." (Papert, 1988, p. 13)

I asked Papert about this when I interviewed him, and he was quite clear. My question was: 'couldn't someone have tried something like back-propagation earlier?' Papert (interview) replied as follows:

> "It wouldn't have made any difference. I don't believe the story that this field [neural networks] took off because of back-propagation. Everybody knew hill-climbing, everybody knew that you can make these systems. The science was around for years, but nobody was paying any attention to it. It was in the culture . . . Clearly, if someone had wanted to work on back-propagation [in the 1960s or 1970s], he wouldn't have gotten much funding . . . [But on the other hand,] if you look in the PDP book [Rumelhart, Hinton, & Williams, 1986] the experiments they did are computationally very tiny, you can run them in your 'PC,' or in your 'Apple.' Anybody could have done them without much funding even in the 1960s."[91]

Papert's words can be understood within the rhetoric of the (reopened) controversy of the late 1980s (this controversy is examined in section 5.2). Papert was trying to create a truism about BP.[92] 'Hill climbing' (equivalent to gradient descent) was 'in the culture' for years, but 'nobody was paying any attention to it.' This may be interpreted as implying that a technique like BP was something obvious ('even in the 1960s'). But even though hill

---

[91] A related comment by Minsky and Papert (1988, pp. 260-261) is: "In the early days of cybernetics, everybody understood that hill-climbing was always available for working on easy problems, but that it almost always became impractical for problems of larger sizes and complexities."

[92] This debating tactic was studied by Star (1989a, pp. 135-137).

climbing methods were known in the 1960s, it was seen in chapter three that something like BP was not obvious at all to neural network researchers of the early 1960s (they tried to solve the problem of training multilayer networks without much success).

Yan le Cun and Robert Hecht-Nielsen pointed out recently that techniques of some similarity with BP were used in control theory as early as the 1950s.

> "In fact, back-propagation is little more than an extremely judicious application of the chain rule and gradient descent . . . Some of the applications and algorithms described in the optimal control literature so closely resemble back-propagation that one could credit Pontryagin (among others) for its discovery [in the late 1950s] . . . From a historical point of view, back-propagation had been used in the field of optimal control long before its application to connectionist systems has been (independently) proposed." (le Cun 1988, pp. 21, 22, and 27)

> "A mathematically similar [to BP] recursive control algorithm was presented by Arthur Bryson and Yu-Chi Ho (1969/1975) in 1969. The primary learning law used can be shown to follow from the Robbibs/Monro technique introduced in 1951(Robbins & Monro 1951; White 1989). The earliest incarnation of backpropagation has probably not yet been found." (Hecht-Nielsen, 1991, pp. 124-125)

But looking for similarities with BP within control engineering for the sake of it is of no use in analysing the development of BP, unless someone uses those ideas from control theory as a resource in developing a neural network algorithm.

In the early 1970s Paul Werbos (at the time carrying out PhD research in applied mathematics at Harvard University) considered the idea of applying steepest descent techniques *plus* 'dynamic feedback' (a technique that he developed) to neural network-like problems. In a multilayer neural network there are (at least) two layers of adjustable connections. The error made by the units in the output layer is easy to calculate. It is just the

difference between the actual output and the desired output for those units. The main problem for minimising a total error function is to calculate the contributions of the internal (hidden) units of the system to that error. This has to be known in order to modify the connections from input units to hidden units. The main problem is therefore to calculate the derivatives of the error with respect to the outputs of the hidden units. Werbos (1974) developed a technique which he called 'dynamic feedback' as a solution to that problem. The idea was to propagate information backwards along the network, so that the derivatives of the error with respect to the intermediate units could be calculated (the idea of propagating error signals backwards will be described in more detail in the section 5.2). Werbos acknowledged that ideas of some similarity with his 'dynamic feedback' had been used in control theory earlier.

> "Werbos (1974) also cited related work in control theory, which also used backwards flows of information to identify systems, albeit in a different way. [Werbos'] formulation [of 'dynamic feedback'] could have been derived as an extension of control theory, but I found it easier simply to prove . . . [it] directly . . . The problem of 'adapting weights' in a neural network is just a special case of the problem of estimating the parameters of a general functional model. The use of square error and steepest descent in estimating a model had been established decades before; therefore, the novel feature of . . . [my formulation of BP] was the use of dynamic feedback in combination with those two components." (Werbos, 1988, p. 341)

In the rest of this section I examine Paul Werbos' (unsuccessful) attempts of applying his 'dynamic feedback' gradient descent technique to neural networks. It was said earlier that gradient descent methods were considered of little interest in AI research in the 1960s and 1970s. This means that, if someone had tried to apply a technique similar to BP to AI-like problems, he or she would have found considerable resistance within the AI community. Paul Werbos (1988) claims that something like this did actually happen to him in the 1970s:

> "A nonlinear version [of back-propagation], essentially equivalent to the generalized delta rule [Rumelhart, Hinton, & Williams' (1986) algorithm], was proposed in various documents circulated in 1971 and 1972. At that time, applications to artificial neural networks were not considered interesting or acceptable to much of the scientific community. Therefore, the method was generalized to permit applications to more conventional forecasting applications (Werbos, 1974)." (p. 341)

I asked Werbos about his work on dynamic feedback in the 1970s. Werbos said that his idea of applying his dynamic feedback method to neural network-like problems was resisted by important people in the scientific community, and that it was not popular at the time. He believes that the difficulties that he had throughout his PhD research at Harvard University were in part related to that. The members of his thesis committee — Werbos said — had doubts about the validity of the early versions of his work. He was told to talk to someone with enough expertise and credibility. This is when, in 1970 or 1971, Werbos went to talk to Marvin Minsky.

> "The response of the Harvard thesis committee was: 'we don't know what to make of this, this is too complicated. You have to prove it to us, and you have to speak to someone reputable'. That's when I spoke to Marvin Minsky . . . I remember going to Minsky at one point saying: 'I have a new model of intelligence.' I gave him some papers. It included back-propagation as a part, only as a part of that. He had a very irascible sense of humour, and said: 'you've been spending all this time, and this is all you come out with. It's not very promising, I don't want you to be working with us at MIT, because this is not promising'. I said: 'look, neurons operate this way'. And he said: 'every neural modeller in the business knows that it follows McCulloch and Pitts' [binary threshold function]. I said: 'yes, the modellers will tell you that, but look at the textbooks where they show you the firing patterns, it may be time sequenced, but it's clearly varying on a whole continuum. So you can get a different model of the neuron that lets you to do derivatives, and lets you to make these things work, and that overthrows what you did in your 'Perceptrons'

> [Minsky & Papert, 1969] book'. Do you know how enthusiastic Minsky was about that? In 1970 or 1971 I presented it to Minsky, and that was his reaction. I think part of it was that I was saying: 'this is a way of getting around your conclusions in the 'Perceptrons' book,' and he wasn't very interested in that kind of a thing." (Werbos, interview)

Minsky told me about Werbos accidentally, without having been asked or told anything about him, in replying to a question about funding for neural networks in the early 1960s. When I told him: 'There are people who say that neural network researchers were trying to get funding in the early 1960s and that they could not, that DARPA would not fund them,' Minsky replied the following:

> "I don't know what they would have done with the money. The story of DARPA and all that is just a myth. The problem is that there were no good ideas. The modern idea of back-propagation could be an old idea. There was someone . . . [trying to remember]. *Question:* Paul Werbos? *Answer*: That's it! [excited]. But, you see, it's not a good discovery. It's alright, but it takes typically 100,000 repetitions. It converges slowly, and it cannot learn anything difficult. Certainly, in 1970 the computers were perhaps too slow and expensive to do it. I know that Werbos thought of that idea. It's certainly trivial. The idea is how you do gradient descent. I didn't consider it practical. *Question:* Because of the computational costs? *Answer:* Yes, but also, with artificial intelligence, we had the experience that when you make a process like that you usually get stuck at a local minimum. We still don't have any theory of what range of problems they work well for." (Minsky, interview)

Thus Minsky recalled the meeting he had with Werbos. At that time, Minsky thought that the idea was not useful enough (he still has many doubts about it nowadays, as I will show in section 5.2). But let me come back to the Werbos story. As time went by (around 1972), and pressure from the Harvard thesis committee was rising, he wrote a simpler and clearer paper about the back-propagation idea. According to Werbos, in that paper he developed the idea of dynamic feedback in the context of multilayer perceptrons.

> "So I pulled off a small piece and said: 'look, I can use this back-propagation part to do pattern recognition in a multilayer perceptron'. I wrote a 20 page paper on how to do this, really straightforward and clean. That was 1972. I can still remember very vividly a good scientist from Harvard University, whose work I strongly respect, saying: 'well, now we understand this. This is all very straightforward. I understand exactly what you want to do, it's clear, it will work. But, you know, this is enough meat for a seminar paper now, this is still not important enough, it isn't good enough to qualify for a Harvard PhD thesis. We can't graduate you on this'. I think that part of the reason why he said this is that he was responding to pressure from some of these peers who didn't like the whole area."(Werbos, interview)

Karl Deutsch (at the time president of the International Political Science Association) suggested that Werbos' technique could be applied to a political science example, and that was accepted by the thesis committee.[93] Thus Werbos did not apply his dynamic feedback algorithm to neural networks in his thesis, although he made some comments about the possibility of doing so.[94]

Later Werbos made further attempts of applying his technique to multilayer neural networks. In 1981 he was working at the 'Center for Computation, Economics, and Statistics' of MIT under the direction of mathematician Charles Smith. Werbos claims

---

93 "Karl Deutsch had written a book called 'Nerves of Government' [Deutsch, 1963]. He was very interested on how the neural network view might be seen as a paradigm for political organisation" (Werbos, interview). "The first actual application of backpropagation was in estimating time-series models used to predict nationalism and social communications, developed by Prof. Karl Deutsch" (Werbos, 1988, 342).

94 " 'Dynamic feedback' is essentially a technique for calculating derivatives inexpensively, for use with the classic method of steepest descent . . . We discuss how our experience here with steepest descent has led to new ways of adjusting the 'arbitrary convergence weights' of steepest descent; these methods speeded up the process of convergence by a large factor . . . We also point out that the algorithms of chapter 2 [the 'dynamic feedback' algorithms], taken as part of 'cybernetics', have a direct value as paradigms, to help us understand the requirements of the complex information-processing problems faced by human societies and by human brains" (Werbos, 1974, pp. xv-xvi). "The mathematics of back-propagation given there do not elaborate on neural nets, although I made sure that I had a chapter which talked about it, and I did discuss neural networks in there. I gave examples in chapter 2 [Werbos, 1974] which are still useful in the neural net profession today" (Werbos, interview).

that he applied to Smith for a project for applying his dynamic feedback method to neural networks, but his application was unsuccessful. Werbos says that Smith afterwards was among the people who funded the researchers of the PDP group for doing 'the same thing'.

> "I gave him [Charles Smith] a little flow chart saying: 'here are multilayer perceptrons, here are derivatives, you can combine them. Furthermore, here is a paper, and I want to do it.' . . . Now, Charlie Smith looked at me, and said: 'this is a workable idea, it does show that we can do something, but you are not the right person, you are a civil servant'. I said: 'But what do you mean? I've only been in the government for two years, you know, and I figured this thing out, and here it is!' . . . I wasn't a member of the right social elite. So then he went out to the System Development Foundation [Menlo Park, California], and was among the people who financed the PDP group's work. He is acknowledged in the beginning of the PDP book for his prominent role in this business."(Werbos, interview)

The paper mentioned by Werbos in this quotation (Werbos, 1982) was presented at the 1981 International Federation for Information Processing (IFIP) conference in New York. In that paper there were explicit allusions to the possible application of Werbos' algorithm to neural network research.[95] C. Smith was the official reviewing officer for that paper, and he did the review before leaving MIT to go to the Systems Development Foundation (Menlo Park, California) (Werbos, interview). In short, Werbos (interview) claims that that the idea of using BP in neural networks originated from him:

> "I am not accusing anyone of plagiarism, but on the other hand I do believe that, causally, I originated the idea, and it spread from me, maybe not always in the form of published papers, but I do believe that the idea did spread from me to

---

95 Here is a quotation from the paper: ". . . Physical networks, made up of units operating in parallel . . . To optimize such a system . . . it is essential to know the derivatives of the desired performance measure with respect to all parameters in the system; for this to be feasible, it is essential to use a method such as the generalized backwards method which does not multiply the cost of getting the derivatives far beyond the cost of exercising the system." (Werbos, 1982, p. 765)

> the relevant places . . . One example of this sticks out in my mind. I remember once (a rare event) going to a party in the Cambridge area, after I had begun my thesis on back-propagation. One of the people, whom I had never met, mentioned that he had heard there was a thing called 'continuous feedback' which would allow you to calculate all the derivatives of a complicated system in just a single pass, and that someone was writing a thesis on it, and you could use it in AI. Once the basic work is done, it is easy for that kind of simplification to spread by word of mouth; this is kind of unwritten, undocumented communication can play a strong role in what happens in science."

Rumelhart (interview) denied firmly Werbos' allegation:

> "Had I known about Werbos' work, I would have been happy to list his name, but we didn't. As far as I know his work was entirely hidden, and nobody knew about it. Werbos is probably sorry that his work didn't have more impact. I'm sure he is, but I think the fact is it didn't. I had no idea of the man. I never heard of the man until well after we published the back-propagation work."

Werbos' view about the unwritten and informal influence of his work on dynamic feedback on Rumelhart, Hinton, and Williams' (1986) work on BP will have to remain as an allegation. In a social activity such as science, the interactions and relationships among the actors are of multiple kinds and go in many different directions. These interactions are very complex, rich, and varied. Furthermore, a good deal of the knowledge used, transmitted, and transformed in the interactions among scientists is tacit, and this makes priority disputes bitter and difficult to solve.

But the concept of 'discovery' which is implied when talking about priority disputes (and by some of Werbos' comments) is not useful within the 'controversy/enrolment of allies and actants/closure' model of scientific change (see chapter one) which I am using here. The term 'discovery' refers to a single, discrete act, localisable in time and space. Robert K. Merton (1961, p. 356) claimed that all scientific discoveries are

'multiples.'[96] But 'multiple discoveries' are still discoveries. The notion of discovering something belongs to a contemplative, passive, realist (and inadequate) model of science (the underlying reality, phenomenon, or procedure is suddenly 'uncovered'). Within the interpretative framework used here (see chapter one), the 'discovery' or 'non-discovery' of a new phenomenon or technique is the outcome of the controversy about the validity of that phenomenon or technique.[97] Harry Collins (1985, p. 89) suggested that:

> "Where there is disagreement about what counts as a competently performed experiment, the ensuing debate is coextensive with the debate about what the proper outcome of the experiment is. The closure of debate about the meaning of competence is the 'discovery' or 'non-discovery' of a new phenomenon."

So the only acceptable meaning of 'discovery' is so far away from what is usually understood by 'discovery' that it is better not to use this term at all. The idea of priority disputes is also of limited value.[98] Of course, there is an obvious sense in which being recognised and credited is important as a source of

---

[96] Merton (1973, part 4) studied scientific discoveries. He claimed that all discoveries in science are in principle multiples: ". . . Far from being odd or curious or remarkable, the pattern of independent multiple discoveries in science is in principle the dominant pattern rather than a subsidiary one. It is the singletons — discoveries made only once in the history of science — that are residual cases, requiring special explanation. Put even more sharply, the hypothesis states that all scientific discoveries are in principle multiples, including those that on the surface appear to be singletons" (Merton 1961, p. 356).

[97] Barnes (1982, p. 45) pointed out that a 'discovery' is the outcome of a social process of validation, and therefore it cannot be a single, discrete event: " 'Discovery' is a social category of approbation denoting the validated status of that to which it refers . . . To say that something is a discovery is to describe it as the outcome of a procedure which at once records and validates it. But it is also . . . to imply that the procedure in question is encompassed within a single act or event. Hence the use of the term 'discovery' implies that validation can be accomplished as one event, truly a sign of an inadequate theory of knowledge."

[98] Priority disputes were often studied in the 'classical' sociology of science (e.g. Merton, 1973; Hagstrom, 1965). W. Hagstrom (1965) reported a case quite similar to Werbos': "In science, the failure to recognize discovery may give rise . . . to strong antagonisms and, at times, to intense controversy . . . [Hagstrom gives the example of an experimental physicist] . . . Something like this [i.e. failure to recognise his accomplishments] had happened earlier in his career, when a grant he had requested was rejected, and shortly afterward someone else had become famous for doing essentially what he had proposed to do" (p. 14-15).

resources for carrying out further research. More interestingly, in Latour's (1987) theory of fact building as a process of enrolling and controlling others (so that they use one's claim and as a consequence this claim becomes more of a 'fact' or a 'black box') credit is important in the sense that a fact is built if people use it as it is, that is if they do not transform it into something else or *someone else's*.

But Werbos was far from being able to push the construction of the BP 'black box' forward. He had a resource (the dynamic feedback algorithm) that he apparently intended to use in neural networks. He deserves credit for having developed that algorithm, but this is not the important question here. The important issue here is that the resistance that Werbos found in the 1970s to the *very idea of back-propagation* was too much for him. He was isolated, and was unsuccessful in enlisting allies and resources to develop his idea. Werbos' technique could have been used as an argument in favour neural networks in the 1970s. The Minsky-Werbos episode is especially significant in this respect. Werbos claimed (in front of Minsky) that he had a solution for the problem of training multilayer networks, and Minsky, as Werbos put it, 'was not interested in that.' Minsky was not interested in revising the conclusions of the closure of the perceptron controversy. Again, the person is not an issue here. In that episode Minsky represented the anti-neural network position. And what is important is that this position remained unchallenged in the years from the closure of the perceptron controversy to the mid- and late 1980s. In section 5.2 I will show that Minsky and Papert were not interested in the revision of the closure of the perceptron controversy in the late 1980s either, but then they were just not powerful enough to stop it.

In this section I have shown that Minsky and Papert (1969) helped construct a reverse salient for neural network research. After their study, training multilayer networks remained as the most important challenge for future neural network researchers. Researchers tried to solve that problem before Rumelhart and

colleagues developed their back-propagation technique in the mid-1980s. In particular, I have studied Werbos' unsuccessful attempts to apply to neural networks a technique similar to back-propagation that he developed in the 1970s. I have concluded that Werbos was not powerful enough to overcome the resistance that he found to the very idea of using something like back-propagation in AI research. In the coming section I analyse Rumelhart and colleagues' back-propagation algorithm in detail.

5.2 Back-propagation: learning in multilayer perceptrons

In this section I look at the back-propagation learning technique for multilayer neural networks. I argue that the development of back-propagation in the mid-1980s was very important for the reopening of the neural network controversy. I also discuss certain aspects of the reopening of the controversy. In particular, I look at Minsky and Papert's (1988) reaction to back-propagation. I show that Minsky and Papert's (1988) renewed criticism of neural networks did not have a decisive effect in the late 1980s (contrary to what had happened in the earlier perceptron controversy). I conclude by pointing out at the interest created by the development of back-propagation.

In the mid-1980s, researchers belonging to the PDP group were more successful than Paul Werbos had been before in developing a back-propagation (BP) neural network. David Rumelhart, Geoffrey Hinton, and Ronald Williams' (1986) BP learning algorithm soon became one of the most successful techniques in neural network research. In the late 1980s BP was widely used and applied to a great variety of problems, and by now it has become an established subarea of research in neural computing. The success of BP was a very important element for the reopening of the perceptron controversy. It was, as it were, the last 'push' that brought about that reopening. After the development of BP, Minsky and Papert's (1969) study was no longer seen as the 'last word' in the perceptron controversy. Minsky and Papert's (1988) quick response (to be studied later in this section) confirms this. Of course, the success of BP has to be understood within the broader context of the re-emergence of neural networks in the 1980s. Some aspects of this emergence process have already been discussed in sections 4.1 and 4.2,

including the PDP group, Hopfield's work, and the Boltzmann Machine. In later sections I will examine other aspects of that re-emergence. In this section I examine the BP network and Minsky and Papert's reaction to it.

At the same time as Rumelhart, Hinton, and Williams (1986) developed their BP network (around the mid-1980s) two other researchers — Parker (1985) in the United States and le Cun (1985) in France — developed similar techniques.[99] There have been no open priority disputes between le Cun, Parker, and Rumelhart and his colleagues, and it is widely accepted that the success of introducing BP in neural computing was due to Rumelhart, Hinton, and Williams. Le Cun (1988, p. 27) recognised this:

> "From a historical point of view, back-propagation had been used in the field of optimal control long before its application to connectionist systems has been (independently) proposed. Nevertheless, the interpretation of back-propagation in the context of connectionist systems, as well as most related concepts are recent, and the historical and scientific importance of (Rumelhart, Hinton, & Williams, 1986) should not be overlooked. The concepts are new, if not the algorithm."

Parker's, le Cun's, and Rumelhart and colleagues' efforts appear to have been carried out independently.

> "We later [after developing the back-propagation algorithm] found that in fact, also I guess as early as 1982, pretty much the same time I guess, David Parker had been working on a similar idea. We also found that Yann le Cun had been working on a similar scheme, although I think that Parker's idea is more similar than le Cun's scheme . . . Some years later we learned that some idea like this had also been proposed by Paul Werbos in the mid-1970s, although it had been totally hidden as far as I know." (Rumelhart, interview)

---

99 For an interview with Parker, see (Swaine, 1989).

Now let me describe the details of Rumelhart, Hinton, and Williams' (1986) BP network. Figure 5-1 shows the architecture of the neural network studied by Rumelhart and colleagues. It is a feedforward, perceptron-like network — that is why it is sometimes called 'multilayer perceptron.' The network is divided into layers, and the units are connected in a feedforward way (in the direction input→ output). The units in one layer (e.g. the hidden layer) are fully connected to the units in the following layer (e.g. the output layer). Thus there are important differences in architecture between networks with symmetric connections such as the Boltzmann Machine (BM) and BP feedforward networks. BP networks are much closer to the architecture of Widrow's madaline and Rosenblatt's perceptron than the BM. But, unlike in Rosenblatt's perceptron, in a multilayer network all the connections have modifiable values. The BP network of figure 5-1 is very similar to some networks studied by Widrow and his colleagues in the 1960s (see section 3.2). However, there are important differences too, and these differences were very important for the development the BP algorithm.

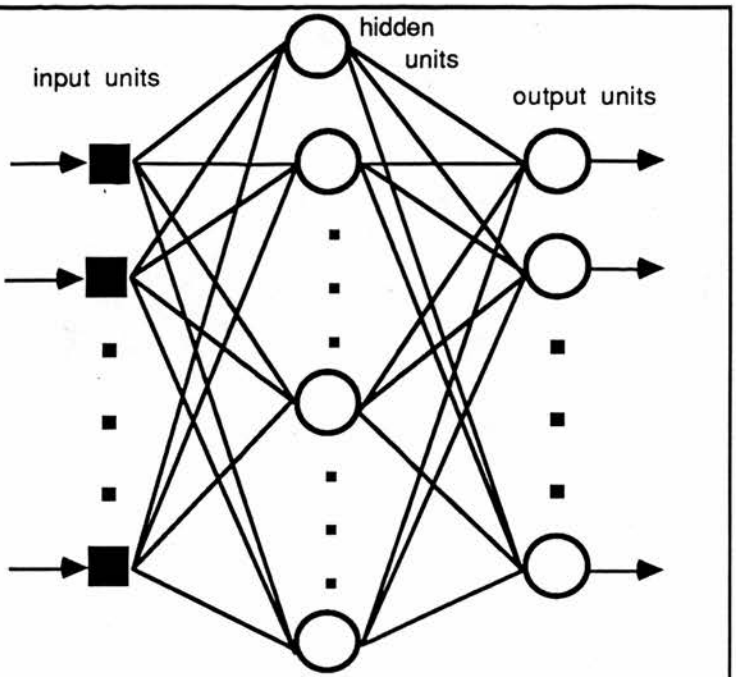


figure 5-1: multilayer perceptron

In the two-layer network of figure 5-1 the input units simply distribute the input activation to the following layer. They do not perform any summation or thresholding (they are represented by small black squares the figure). Units in both the hidden layer and the output layer are similar to the usual processing units in early neural networks. However, there is one important difference between the processing units used by early researchers such as Rosenblatt and Widrow and the ones used by Rumelhart and colleagues. This difference was crucial for the solution given by Rumelhart and his colleagues to the problem of training multilayer feedforward networks. Rumelhart and his colleagues used a continuous, differentiable, sigmoid threshold function instead of the step function used by early neural network researchers. Figure 5-2 below represents a step function, and figure 5-3 is an approximation to a sigmoid function. Units with sigmoid activation function had been used

used in Hinton and colleagues' BM, and in earlier neural networks by researchers like Grossberg (see section 4.2).

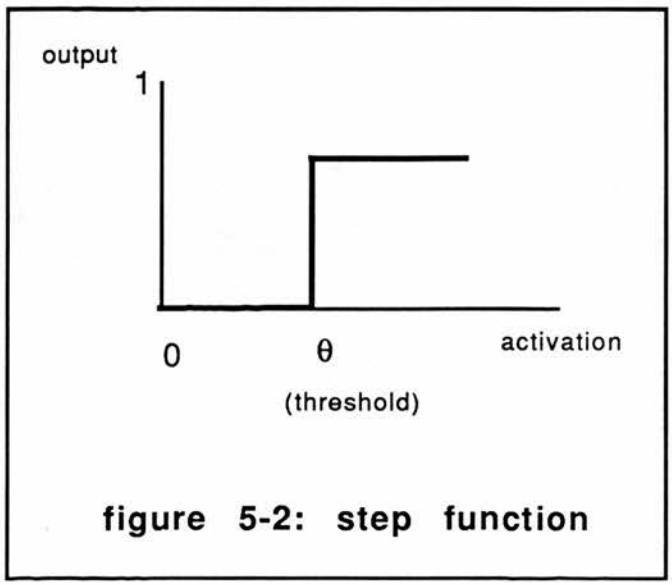


figure 5-2: step function

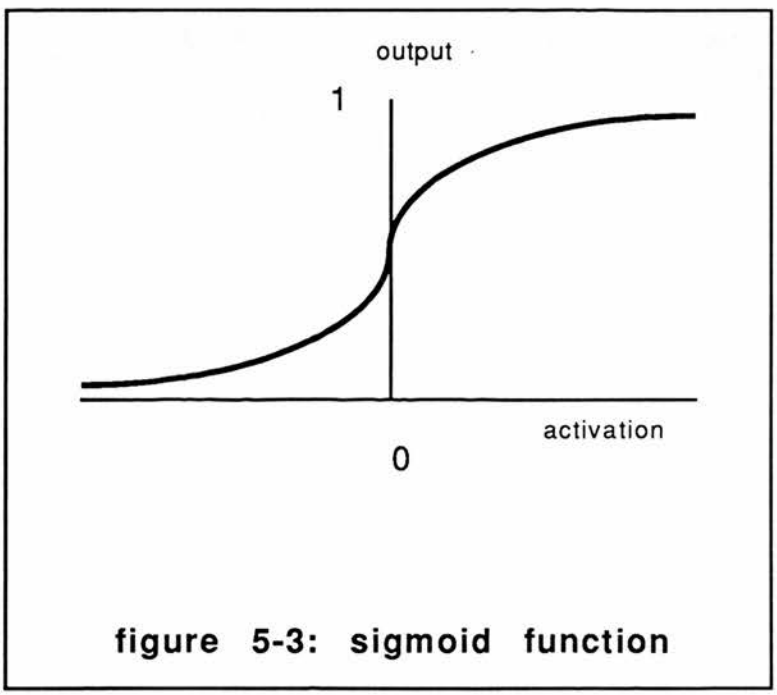


figure 5-3: sigmoid function

The input activation ('net_{pj}'in Rumelhart and colleagues' symbols) to a processing unit (unit j) in a BP network when a pattern (pattern p) is presented to the system is then:

$$net_{pj}=\sum_{i} w_{ji}\ o_{pi}$$

This summation is the usual one in processing units of neural networks. 'w_{ji}' are the values of the connections (from units i) coming to unit j, and 'o_{pi}' represents the outputs of of those units i when pattern p is presented. (It is important to note that o_{pi} represents the inputs to unit j. It should not be confused with o_{pj}, the output of unit j.)

The weighted sum is then thresholded. The thresholding function used by Rumelhart and his colleagues was a continuous, differentiable sigmoid function. The output (o_{pj}) of a processing unit (of the hidden or output layer) when pattern p is presented is then:

$$o_{pj} = f(net_{pj}) = \frac{1}{1+e^{-\kappa\ net_{pj}}}$$

'κ' is a positive constant which controls the 'spread' of the function (Beale & Jackson, 1990, p. 72). The range of o_{pj} is: $0 \leq o_{pj} \leq 1$. The reason why Rumelhart and his colleagues used the above function is that its derivative, in symbols $f'_j(o_{pj})$, is a simple function of o_{pj}:

$$f'_j(o_{pj}) = \kappa\ o_{pj}\ (1 - o_{pj})$$

This derivative intervenes in the weight adjustment process and has got some interesting properties. It reaches its maximum at $o_{pj}=0.5$, and approaches its minimum as o_{pj} approaches 0 or 1, and therefore it ensures that weight changes will be biggest for those units which are near their midrange activation (0.5), and thus not yet committed to being either on or off (Rumelhart, Hinton, & Williams, 1986, p. 329). This derivative also makes sure that the weights remain within a bounded range. If the activation o_{pj} becomes too large, then the derivative $f'(o_{pj})$ goes to zero, and this ensures that the weights stop growing (Carpenter, 1989, p. 247).

The objective of the BP weight-modification algorithm is to minimise the error made by the network (that is by the output

units of the network) in some classification task. The error measure used by Rumelhart, Hinton, and Williams corresponded to Widrow and Hoff's least mean square (LMS) error. The total error is $E=\Sigma E_p$. The error for input/output pattern p is defined as:

$$E_p = \frac{1}{2} \sum_j (t_{pj} - o_{pj})^2$$

In words, the error for pattern p is proportional to the square of the difference between the actual output (o_{pj}) and the desired output (t_{pj}). Rumelhart, Hinton, and Williams (1986, pp. 322-328) derived their BP learning equations as a generalisation of Widrow and Hoff's weight-modification algorithm using the 'chain rule' for differentiation. (Rumelhart and colleagues called Widrow and Hoff's algorithm the delta rule, and that is why the BP algorithm is sometimes referred to as the 'generalised delta rule.')

Rumelhart, Hinton, and Williams (ibid., p. 322) defined Widrow and Hoff's weight modification rule as:

$$\Delta_p w_{ji} = \eta (t_{pj} - o_{pj}) i_{pi} = \eta \delta_{pj} i_{pi}$$

$\Delta_p w_{ji}$ is the weight change in the connection from unit i to unit j after presentation of pattern p. o_{pj} is the actual output of unit j after presentation of pattern p. t_{pj} is the desired output for that unit (for pattern p). i_{pi} is the input i to unit j when pattern p is presented (i.e. the i-th value of the input pattern). (i_{pi} could also have been written o_{pi}. In the equation above i_{pi} means input i to unit j, but the input to that unit is also the output of another unit (unit i), that is o_{pi}.) δ_{pj} is the error made by unit j.

What one needs to know in order to adjust the weights in a multilayer BP network is the error made by each unit. The error made by the units in the output layer is easy to calculate. It is the difference between the actual output pattern produced by the network and the desired output pattern. But it is not obvious how to calculate the error made by each of the units in the hidden

layer (and this is necessary in order to be able to adjust the connections between input units and hidden units). The intuitive idea of back-propagation is that the error made by a hidden unit (unit j) should depend on the errors made by the output units (units k) to which that unit (unit j) is connected. These errors are back-propagated, so that the weights between input units and hidden units can then be adjusted. In a BP network each output unit demands from the hidden units exactly what it needs, and the hidden units try to accommodate the conflicting demands.

The term 'back-propagation,' as well as the intuitive idea of making the error of a hidden unit proportional to the error made by the output units to which that hidden unit is connected, goes back to Frank Rosenblatt himself (see Rosenblatt, 1962a, ch. 13).

> ". . . Considerable improvement in performance might be obtained if the values of the S [sensory units] to A [association units] connections could somehow be optimized by a learning process . . . The difficulty is that whereas R*, the desired response, is postulated at the outset, the desired state of the A-unit is unknown . . . The 'back-propagating error correction procedure' . . . takes its cue from the error of the R-units [response or output units], propagating corrections [Rosenblatt probably means 'propagating errors'] back towards the sensory end of the network if it fails to make a satisfactory correction quickly at the response end . . . At present, no quantitative theory of the performance of systems with variable S-A connections is available" (Rosenblatt, 1962a, pp. 287-298).

Note the last sentence in this quotation by Rosenblatt, indicating that no algorithm had been developed until then to carry out the idea of back-propagation.

In section 5.2 it was said that methods of a certain similarity with BP were used in control theory in the 1960s, and that Paul Werbos developed a technique equivalent to BP — his 'dynamic feedback' algorithm — and suggested its application to multilayer neural networks unsuccessfully in the 1970s and early 1980s.

But let me come back to Rumelhart and colleagues' BP technique. After the weights and thresholds of the network have been initialised (i.e. set to small, random values), an input pattern is presented, and the network produces an output (the actual output). The general equation used by Rumelhart, Hinton, and Williams to adjust the weights of their network was the same as Widrow and Hoff's (as it was said earlier). That is:

$$w_{ji}(\tau+1) = w_{ji}(\tau) + \eta\ \delta_{pj}\ o_{pi}$$

BP1

τ represents time, and η is a constant which determines the rate of weight modification. δ_{pj} is the error made by unit j after presentation of pattern p. It can be seen in BP1 that, in order to adjust the weights of the network, the error made by each unit (δ_{pj}) must first be calculated. For the output units Rumelhart and colleagues derived the following equation for calculating the error:

$$\delta_{pj} = (t_{pj} - o_{pj})\ f'_j(o_{pj}) \quad \text{,that is:}$$

$$\delta_{pj} = \kappa\ o_{pj}\ (1 - o_{pj})\ (t_{pj} - o_{pj})$$

BP2

The error made by a hidden unit (unit j) depends both on the error of the output units (units k) to which that hidden unit is connected (δ_{pk}) and on the values of the connections from unit j to units k (w_{kj}).

$$\delta_{pj} = f'_j(o_{pj}) \left(\sum_k \delta_{pk}\ w_{kj} \right) \quad \text{,that is:}$$

$$\delta_{pj} = \kappa\ o_{pj}\ (1 - o_{pj}) \left(\sum_k \delta_{pk}\ w_{kj} \right)$$

BP3

Because the hidden units have no access to information about the error made by the output units to which they are connected (δ_{pk}), this information is back-propagated throughout the w_{kj} connections (and it is multiplied by the value of those connections). This backward pass has the same computational complexity as the forward pass of activity in the network (Rumelhart, Hinton, & Williams, 1986, p. 327).

A (forward-backward) learning cycle in a BP network can be summarised as follows. A pattern p is presented, activity propagates forward throughout the units, and the network produces an output. This output is compared to the desired output, and the error made by the output units is calculated according to BP2. Then, before any weight adjustment is made, the backward stage starts. The errors made by the output units are back-propagated (through the 'old' hidden-to-output connections) to the hidden units, so that the error made by each hidden unit can be calculated according to BP3. Now all the connections in the system can be changed according to BP1. If there were more layers of connections, those layers would be adjusted in the same way. It is important to note that the whole backward pass has to be completed before any weight adjustment is made. For variations and further developments of BP, see Hertz et al., 1991, ch. 6).

BP minimises the error made by the system (i.e. the errors made by the output units; the hidden units do not have target values) over a set of patterns. In other words, by adjusting the connections of the system according to the BP technique, the total error measure for a set of input/output patterns is minimised in a gradient descent way.

> "To minimize E [total error] by gradient descent it is necessary to compute the partial derivative of E with respect to each weight in the network. This is simply the sum of the partial derivatives for the input-output cases. For a given case, the partial derivatives of the error with respect to the weight are computed in two passes [the

forward pass and the backward pass]." (Rumelhart, Hinton, & Williams, 1986b, p. 697)

The backward pass was criticised by neural network researcher Stephen Grossberg. Grossberg (1987, pp. 47-50) claimed that the 'weight transport' required in the backward pass of BP had no possible physical interpretation from a 'brain modelling' point of view.[100] However, by and large, the reaction to BP within the emerging neural network community was rather optimistic. In fact, BP was seen by most neural network researchers as the most important result of the mid- and late 1980s. The quotation below by Bernard Widrow is an example in this respect. Widrow also stresses the importance of substituting a sigmoid threshold function for the step function (the typical one in early neural network research in the 1960s).

> "The publication of the backpropagation technique by Rumelhart, Hinton, and Williams (1986) has unquestionably been the most influential development in the field of neural networks during the past decade. In retrospect, the technique seems simple. Nonetheless, largely because early neural network research dealt almost exclusively with hard-limiting non-linearities, the idea never occurred to

---

[100] Grossberg (1987, pp. 47-50) was quite critical about this 'weight transport' issue, and indeed about the whole idea of BP as a 'brain modelling' tool. But one should not forget that Rumelhart and his colleagues did not claim to have produced a tool for modelling real brain activity. Curiously, Carpenter and Grossberg's (1987, 1988) ART (adaptive resonance theory) system has been criticised on similar grounds. In his discussion of BP, Grossberg uses a terminology based on levels. The translation to the terminology being used in this dissertation would be the following: F1 is the input level (or layer), F_2 is the hidden unit level, and F_3 is the (actual) output level. Grossberg uses further levels to represent the backward pass in BP: F_4 is the level where error signals for the output units are computed, and F_5 is the level where the error signals of the hidden units are computed. Thus according to Grossberg weight are transported from the $F_2 \rightarrow F_3$ to the $F_4 \rightarrow F_5$ connections. Here is his view: "Back propagation proceeds as follows. The weights computed in the bottom-up $F_2 \rightarrow F_3$ pathways are *transported* to the top-down $F_4 \rightarrow F_5$ pathways . . . Such a physical transport of weights has no plausible physical interpretation. The weights in the $F_2 \rightarrow F_3$ pathways must be computed *within* these pathways in order to multiply signals from F_2 to F_3. These weights cannot also exist within the pathways from F_4 to F_5 in order to multiply signals from F_4 to F_5 without being physically transported from ($F_2 \rightarrow F_3$) to ($F_4 \rightarrow F_5$) pathways, thereby violating basic properties of locality . . . The BP model is thus not a model of a brain process" (Grossberg, 1987, pp. 49-50).

neural network researchers throughout the 1960s." (Widrow & Lehr, 1990, p. 1433)

In neural network research BP is seen by most as a successful solution to the reverse salient of training multilayer perceptrons. A key change for the development of BP — and therefore in the reformulation of the reverse salient of learning in multilayer networks as a 'critical problem' (using Hughes' terms) — was the substitution of a smooth, continuously differentiable function for the step function typical of early neural networks. It was said in section 3.3 that for Hughes (1983) the key to innovation in technology is the redefinition of reverse salients as critical (solvable) problems. In a way, the development of BP can be interpreted within this scheme. The following comment by Terrence Sejnowski, who developed the first important application of BP (Sejnowski & Rosenberg, 1986) can be understood within the reverse salient/critical problem scheme:

> "It is really interesting to reread Nilsson's (1965) book [on early neural networks] because you can see the exact assumptions, often they are very small assumptions, having to do with how you define the input-output function.[101] They used discontinuous step functions. Now we use sigmoids, which have a continuous transition. It may seem like a very small change, because they are very similar functions in terms of their overall nonlinearity, but mathematically it's like night to day. A function like that made it possible learning in multilayer networks. The thing about mathematics is that you can prove beautiful theorems, but you have to make assumptions. You change one of the assumptions, even the smallest, and then a lot of things will change. In particular, something that you couldn't see beyond, suddenly dissolves, or you find a way of getting around it." (Sejnowski, interview)

A great part of the importance of the BP algorithm is caused by the classification power of multilayer neural networks. This was

---

101 When I interviewed Sejnowski, he was writing an introduction for the new edition of Nilsson's (1965) 'Learning Machines' book (see Nilsson, 1990, pp. vii-xxi).

an attractive property of multilayer systems since the early 1960s. The following general comment from Hawkins' (1961, p. 47) early review of 'self-organising' systems is an example:

> "A number of alternatives [to the problems of single layer networks] are possible . . . The most attractive . . . appears to be multiple-layer logical circuit arrangements, since it is known that any function can thereby be realized."

After Rumelhart and colleagues' results on BP, the classification power of multilayer networks received increasing attention. Classifications realised by neural networks can be represented as decision regions in pattern space. For instance, for a network with two input units and one (binary-valued) output unit (and therefore two weights), the pattern space can be represented by four points (11, 00, 01, 10) in a two dimensional space (spaces of more than three dimensions are very difficult to visualise). The classification of these four inputs can be seen as a straight line which divides the input space into two regions. In section 3.3 it was seen that a network of this kind can realise, for example, the 'and' function.

Multilayer neural networks with two layers of hidden units and three layers of modifiable connections (see figure 5-4) can form any decision region in pattern space, that is they can realise decision regions (classifications) of arbitrary complexity (this complexity being limited by the number of units in the system) (Lippmann, 1987, pp. 15-18; Darpa, 1988, pp. 78-80). In other words, a multilayer network with two layers of hidden units can realise any input/output function (or classification).[102] Remember that the BP weight modification algorithm described earlier for a network with two layers of adjustable connections applies equally to a network with three such layers.

---

102 ". . . Three-layer perceptrons with two hidden layers . . . can form any desired decision region . . . Kolmogorov . . . proved a theorem described in Lorentz (1976) which, in effect, demonstrates that a three-layer network can form any continuous nonlinear function of the inputs" (Darpa, 1988, pp. 79-80).

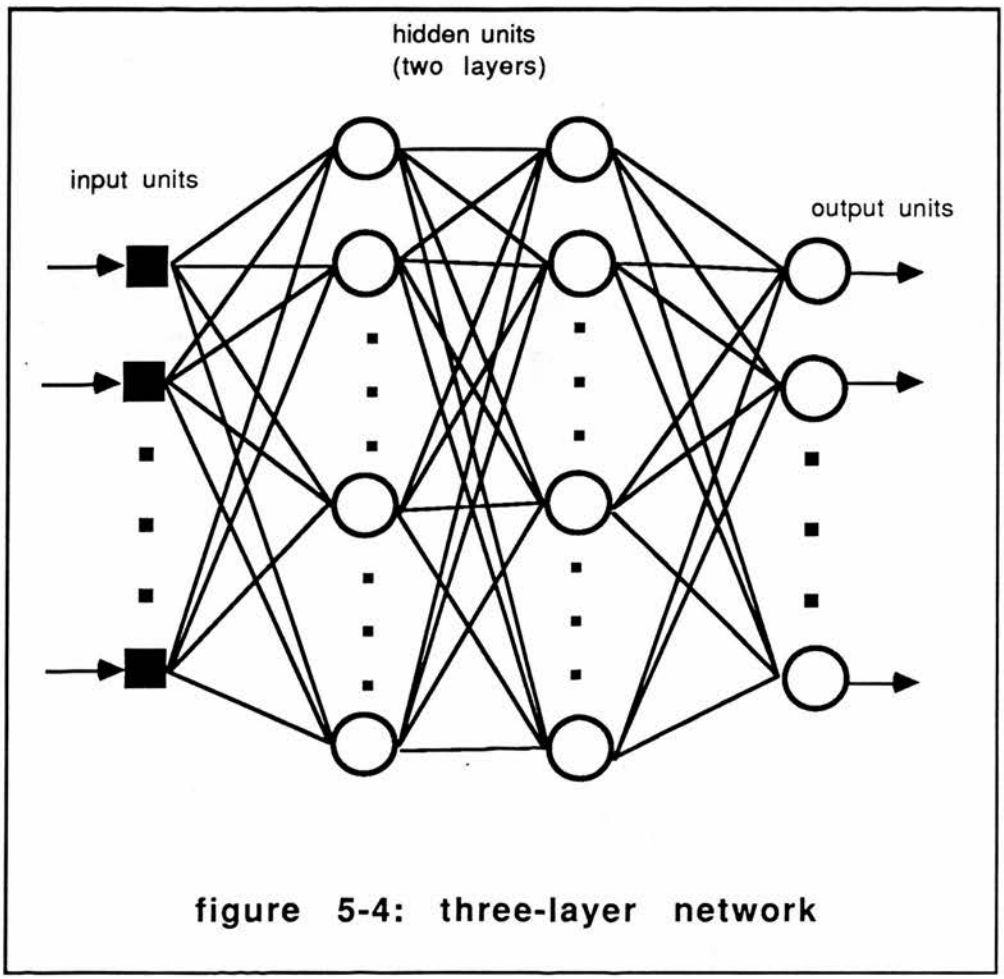


figure 5-4: three-layer network

But, although of great importance, in-principle classification power is only one one aspect of neural computing. Other very important issues include the number of learning cycles required for a classification task, and the number of units required. It will be seen below that Rumelhart and colleagues' claims were pretty much about questions of practice ('empirical' questions, as they put it).

It is important to note that, right from the very beginning, Rumelhart and his colleagues situated their contribution in the context of the perceptron controversy, and in particular in the context of Minsky and Papert's (1969) criticism of single-layer networks and their challenge about multilayer networks. It was seen in section 3.3 that, after analysing the problems of the single-layer perceptron, Minsky and Papert concluded their study

with a (by now rather famous) pessimistic 'intuitive judgement' about the possibility of finding a training algorithm for multilayer networks. This intuitive judgement was quoted in section 3.3, but it is worth reproducing it here too.

> "The perceptron has shown itself worthy of study despite (and even because of!) its severe limitations. It has many features to attract attention: its linearity; its intriguing learning theorem; its clear paradigmatic simplicity as a kind of parallel computation. There is no reason to suppose that any of these virtues carry over to the many-layered version. Nevertheless, we consider it to be an important research problem to elucidate (or reject) our intuitive judgement that the extension is sterile. Perhaps some powerful convergence theorem will be discovered, or some profound reason for the failure to produce an interesting 'learning theorem' for the multilayered machine will be found" (Minsky & Papert, 1969, pp. 231-232).

In section 5.1 it was said that, in a sense, Minsky and Papert can be said to have helped construct the reverse salient of learning in multilayer neural networks. Rumelhart, Hinton, and Williams (1986) claimed that their BP learning algorithm was a successful response to Minsky and Papert's challenge. Even though they were aware that the BP network gets sometimes trapped in local (or false) minima of the error landscape, they claimed that, *in practice*, this was not a significant problem.

> "The problem, as noted by Minsky and Papert, is that whereas there is a very simple guarantied learning rule for all the problems that can be solved without hidden units, namely, the perceptron convergence procedure (or the variation originally due to Widrow and Hoff, which we call the delta rule), there is no equally powerful rule for learning in networks with hidden units . . . The standard delta rule [Widrow's LMS or delta rule algorithm] essentially implements gradient descent in sum-squared error for linear activation functions. In this case, without hidden units, the error surface is shaped like a bowl with only one minimum, so gradient descent is guarantied to find the best set of weights. With hidden units, however, it is not so obvious how to compute the

> derivatives, and the error surface is not concave upwards, so there is the danger of getting stuck in local minima. The main theoretical contribution of this [paper] is to show that there is an efficient way of computing the derivatives. The main empirical contribution is to show that the apparently fatal problem of local minima is irrelevant in a wide variety of learning tasks . . . Although our learning results do not *guarantee* that we can find a solution for all solvable problems, our analysis and results have shown that as a practical matter, the error propagation scheme leads to solutions in virtually every case. In short, we believe that we have answered Minsky and Papert's challenge and *have* found a learning result sufficiently powerful to demonstrate that their pessimism about learning in multilayer machines was misplaced." (Rumelhart, Hinton, & Williams, 1986, pp. 321, 324, and 361)

It was quite clear that, if statements like this were accepted in the AI research community, then this meant that Minsky and Papert's (1969) criticism of early neural networks did no longer apply to more recent work, and that the perceptron controversy had reopened. Minsky and Papert's (1988) quick reaction confirmed this. Rumelhart and colleagues' optimistic evaluation of BP was criticised by Minsky and Papert. In a (limited) sense one could say that history repeated itself: twenty years after their critical study about Rosenblatt's perceptron, Minsky and Papert (1988) criticised Rumelhart and colleagues' claims about their BP learning algorithm. The difference between Minsky and Papert's evaluation of the BP results and Rumelhart and colleagues' original claims was striking. This was of course another case of interpretative flexibility of scientific results.

> "We have the impression that many people in the connectionist community do not understand that this [back-propagation] is merely a particular way to compute a gradient and have assumed instead that back-propagation is a new learning scheme that somehow gets around the basic limitations of hill-climbing . . . Virtually nothing has been proved about the range of problems upon which GD [the generalised delta rule, or BP] works both efficiently and dependably. Indeed, GD can fail to find a solution when one

> exists, so in that narrow sense it could be considered *less* powerful than PC [the perceptron convergence procedure]. In the early years of cybernetics, everybody understood that hill-climbing was always available for working easy problems, but that it almost always became impractical for problems of larger sizes and complexities . . . The situation seems not to have changed much — we have seen no contemporary connectionist publication that casts much new theoretical light on the situation . . . We fear that its [BP's] reputation also stems from unfamiliarity with the manner in which hill-climbing methods deteriorate when confronted with larger-scale problems. In any case, little good can come from statements like 'as a practical matter, GD leads to solutions in virtually every case' or 'GD can, in principle, learn arbitrary functions.' Such pronouncements are not merely technically wrong; more significantly, the pretense that problems do not exist can deflect us from valuable insights that could come from examining things more carefully. As the field of connectionism becomes more mature, the quest for a general solution to all learning problems will evolve into an understanding of which types of learning processes are likely to work on which classes of problems. And this means that, past a certain point, we won't be able to get by with vacuous generalities about hill-climbing. We will really need to know a great deal more about the nature of those surfaces for each specific realm of problems that we want to solve." (Minsky & Papert, 1988, pp. 260-261)

The rhetoric employed by Minsky and Papert in this quotation was a sign of the reopening of the controversy. It included expressions like: 'nothing has been proved;' 'technically wrong;' 'can deflect us from valuable insights;' and 'vacuous generalities.' I asked Minsky about the claim by Rumelhart and colleagues that his 'intuitive judgement' about training multilayer systems was misplaced and that an effective training technique for multilayer perceptrons has now been developed. Minsky (interview) replied as follows:

> "The book [Minsky & Papert, 1969] does say that we don't think that there is an efficient way to make multilayer networks learn. Now, 'efficiently' in the 1960s meant a few thousand trials. Of course now if it does it in a million

> trials it is not so bad. . . There are two issues here. One is that one sense of 'efficiency' has changed. We don't care if it is a million now. The other is that we don't know if the new networks [BP networks] solve any difficult problems . . . When someone demonstrates that a neural network learns some task, that does not mean that a symbolic system cannot do it . . . But I agree that the symbolic approach will have a good deal of trouble if they don't have some fuzzyness."

What 'the book' (Minsky & Papert, 1969) says on training multilayer networks can be seen in the (ibid., pp. 231-232) quotation above, and was discussed in section 3.3. What interests me here (and I showed in section 3.3 and 3.4) is that the book was widely interpreted as showing that learning in multilayer systems was a hopeless problem and that the neural network approach was not worth pursuing. The important thing in the (Minsky, interview) quotation above is that Minsky seems to imply that their (Minsky & Papert, 1969) conclusions about learning in multilayer networks still hold. The concept of efficiency has changed (because of developments in computer technology), but 'we don't know if BP networks solve any difficult problems.' Minsky also tries to make the new BP algorithm more of a truism ('efficiently in the 1960s meant a few thousand trials, but of course now if it does it in a million trials it is not so bad'). He also uses 'bear with us' (Star, 1988a, pp. 137-138) tactics: symbolic systems can get similar results in the near future. Another tactic frequently used by Minsky to defuse the neural computing/symbolic AI opposition is to emphasise the differences within symbolic AI. Minsky (interview) pointed out that he has always been in favour of more 'fuzzy' approaches.[103]

The way in which Rumelhart and his colleagues see the problems of hill-climbing (or gradient descent) techniques is remarkably different from that of Minsky and Papert.

---

[103] Examples of this are Minsky's (1987) 'society of mind' approach, and his differences with McCarthy.

". . . The procedure we have produced [BP] is a gradient descent method and, as such, is bound by all the problems of any hill-climbing procedure — namely, the problem of local maxima or (in our case) minima. Moreover, there is a question of how long it might take a system to learn . . . However, we have carried out many simulations which lead us to be optimistic about the local minima and time questions . . ." (Rumelhart, Hinton, & Williams, 1986, p. 328)

These disagreements in the evaluation of Rumelhart and colleagues' BP technique are another case of interpretative flexibility, that is of divergent views of the same experimental results. What for Rumelhart, Hinton and Williams is a technique 'that leads to solutions in virtually every case' for Minsky and Papert is an impractical technique that deteriorates as problems become larger (and more realistic) than those studied by Rumelhart and colleagues. Different groups of researchers with different goals and interests disagree on the evaluation of the solution given to the reverse salient of learning in multilayer neural networks.

It is interesting to describe an example of this disagreement. Rumelhart, Hinton, and Williams (1986, pp. 334-335) described a small neural network that was able to compute parity. The (mini) network has four input units, four hidden units, and one output unit. Since it has four input units, and the input patterns are vectors of 1s and 0s, there can be up to 16 different input patterns. After presenting these 16 input vectors to the system 2,825 times each, that is after 45,200 input presentation/connection adjustment cycles, the network learned to classify the patterns correctly. For Rumelhart and colleagues this small example showed that BP was a successful learning algorithm for multilayer networks. Minsky and Papert (1988, p. 254) doubted it:

". . . Thus consuming 45,200 trials for the network to learn to compute the parity predicate for only four inputs. Is this a good result or a bad result? We cannot tell without more

knowledge about why the procedure requires so many trials."[104]

Minsky (interview) made the following remark on the parity problem:

> "We don't have any theory of what range of problems they [BP networks] work well on. For example, they don't work on parity, as far as I know, and yet the connectionists say: 'yes, my machine learned to find the exclusive or for six inputs,' or something like that." (Minsky, interview)

There were other points of disagreement between Minsky-Papert and Rumelhart and colleagues. For instance, Minsky and Papert (1988, p. 252) claimed that "multi-layer networks will be no more able to recognize connectedness than are perceptrons." So in the late 1980s problems like parity and connectedness were again a matter of controversy between Minsky and Papert and neural network researchers (Rumelhart and colleagues this time), as they had been at the time of early neural network research.

There are certain parallels between Minsky and Papert's (1988) criticism of BP and their earlier criticism of the perceptron and neural network research. First, Minsky and Papert (1988) criticised the adequacy of BP, and gradient descent methods in general, for AI research, as they had done earlier (see also section 5.1 on the history of BP). And secondly, Minsky and Papert (1988) claimed that problems like parity had not been successfully solved yet. In chapter three it was shown that Minsky and Papert's (1969) criticism was of great importance in

---

104 Papert discussed a similar example, the computation of the exclusive-or function by a neural network. Rumelhart, Hinton, and Williams (1986, pp. 330-331) simulated a system with two input units, one hidden unit, and one output unit. After presenting each input 558 times to the network, that is (as there are four different stimuli) after 2,232 training cycles were completed, the system finally found a solution. "Exor's [exclusive or] learning process consumed 2,232 repetitions of a training cycle; in each repetition the machine was presented with one of the four possible combinations of inputs (one-one, zero-zero, zero-one, one-zero) and a feedback signal to indicate whether it had given the right response ('no' for the first two and 'yes' for the others). Smart or stupid? Should one be more impressed by the fact that the thing 'learned' at all, or by the fact that it learned so slowly and laboriously?" (Papert 1988, pp. 5-6).

the 'crystallisation' (using Harry Collins' term) of the consensus against neural networks in the late 1960s. But Minsky and Papert's (1988) renewed criticism did not have the same 'crystalising' effect the late 1980s. Their 1988 arguments were not seen as decisive against neural networks in the recent re-emergence of neural network research.

The reopening of the perceptron controversy could not be stopped in the late 1980s, and arguments that were seen as decisive in the late 1960s no longer seemed so. In spite of Minsky and Papert's (1988) renewed criticism, many researchers now see gradient descent methods like BP as perfectly adequate in AI research, and Rumelhart and colleagues' BP results are widely seen as an adequate solution to the reverse salient of training multilayer networks.

Of course, I do not mean that Minsky and Papert's (1988) renewed criticism was misplaced or exaggerated (it is not my business to judge it). This criticism, like the 1969 one, created (constructed) problems for neural network research. I said earlier that neural network research has been shaped through controversies with its opponents, and in this sense its opponents have contributed to the field.

BP multilayer networks have become a subarea of research in its own right, with an increasing number of researchers devoting their careers to developing, improving and applying them.[105] One of the earliest important applications of BP was Terrence Sejnowski and Charles Rosenberg's (1986, 1987) NETtalk network, a BP multilayer neural network that learns to pronounce an English text. NETtalk (see figure 5.5) has three layers of units: input, hidden, and output. The input layer has 203 units (7 groups of 29 units), the hidden layer has 80 units, and the output layer has 26 units. The input units are fully connected to the hidden units, and these are also fully connected to the output units.

---

105 For a review on some attempts to minimise the occurrences of local minima when using BP see (Beale & Jackson, 1990, pp. 91-97).

In NETtalk local representation was used in the input layer: the letters were represented locally within each group of input units (26, plus three units to encode punctuation and word boundaries, making a total of 29 units per group). (Anderson and Rosenberg [1988, p. 662] pointed out that Sejnowski and Rosenberg later carried out simulations with distributed input representations, and got comparable results.) The output units represented phonemes in a distributed way. Phonemes were represented in terms of 23 articulatory features, with three additional units to encode stress and syllable boundaries. The output units were the input to a speech synthesiser. The hidden units developed their own 'distributed internal representations' through BP learning, and worked as feature detectors.

Figure 5-5 is an schematic representation of NETtalk. It is important to note that the input to the system consists of seven letters (one for each group of input units), and that the system in figure 5-5 is 'reading' the letter in the fourth (central) square. The other six letters provide a partial context for the decision (Sejnowski and Rosenberg wanted to account for the context-sensitive aspects of English pronunciation). In figure 5-5, the expression 'a cat' is being fed to the input units, and the system is reading the letter 'c' (phoneme /k/).

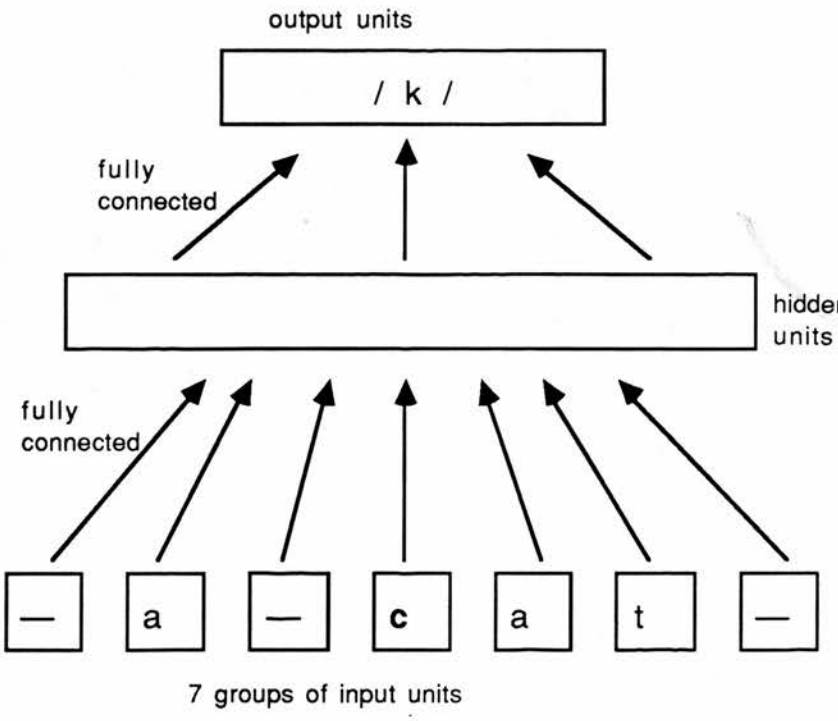


figure 5-5: NETtalk system

Sejnowski and Rosenberg carried out several experiments with their NETtalk system. In one of them, they used a training set consisting of 1024 words. They reported (1986, p. 665) that after 50,000 training cycles (roughly 12 CPU hours of training on a DEC VAX computer, according to Anderson and Rosenfeld [1988, p. 662]), the network's performance reached 95% accuracy. Afterwards, the network's capacity for generalisation was tested by presenting to it a 493 word continuation from the same speaker without training. Sejnowski and Rosenberg (1986, p. 667)reported accuracy of 78% for this case.

Experiments like those carried out with NETtalk caused a good deal of excitement about neural networks in the second half of the 1980s. From the very beginning, NETtalk was compared to DECtalk (from 'Digital Equipment Corporation'), the state of the art, commercially available rule-based expert system for text to

speech synthesis. DECtalk (a result of many years of research) outperformed NETtalk (developed in a summer), but researchers were impressed by the speed of learning of NETtalk, and saw it as a promise of the capabilities of neural networks. DECtalk (based on D. Klatt's work) took about 15 years of research to develop. For neural network researchers NETtalk showed the advantages of neural computing. However, D. Klatt (DARPA, 1988, p. 219) recently 'counter-attacked' by claiming that the (acoustic-phonetic rule based) alphabetic-to-phonetic representation system of DECtalk took only three years to develop. NETtalk was not pursued further towards commercialisation (ibid., 220), but it is a good example of the early — and for neural network researchers promising — applications of networks with BP.

Some of the early applications of BP were summarised in DARPA's (1988) neural network study. According to this study, BP was being used at that time in application areas including pattern recognition and classification, signal processing, and speech recognition. Pattern recognition and classification applications of multilayer BP networks included 'tactical target recognition using radar imagery,' 'discrimination between two different sonar targets,' and 'smart weapons.' Signal processing applications included 'recovery of noise-corrupted or distorted waveforms,' and 'prediction of time series.' Finally, speech applications included 'text-to-speech synthesis' and 'speech recognition.' (see DARPA, 1988, pp. 203-221).

Since then multilayer BP perceptrons are been applied to a great variety of problems. An study of these applications is out of the scope of this dissertation. For a review of some neurobiology-oriented applications, see A. Anderson (1988). Le Cun et al. (1989) applied multilayer BP networks to handwritten post code recognition in the United States. Beale and Jackson (1990, pp. 97-104) reviewed applications in areas such as predicting seat demand in airlines (Airline Marketing Tactician), electro-cardiograph noise filtering, movement of financial markets, bank

loan scoring, aircraft identification, terrain matching for navigation systems, target identification from sonar traces, monitoring level crossings (British Rail), speech processing, recognition and synthesis (British Telecom), and cheque signature verification. Another application,developed by T. Sejnowski and S. Lekhy in 1988, is a neural network which computes curvature (shapes, depth) from shading in an image (see Anderson 1988, p. 657). More recently, Hertz et al. (1991, pp. 134-141) reviewed applications of BP networks in several problems. These include NETtalk-like systems for prediction of secondary structure of proteins (Qian & Sejnowski, 1988) (which allegedly outperformed the best available alternative method), and for hyphenation. Other applications reviewed by Hertz et al. include sonar target recognition, navigation of a car, image compression, signal prediction and forecasting, and backgammon. The 'Neurogammon' system was developed by G. Tesauro and T. Sejnowski in 1988, and a later version of it (Tesauro, 1990) defeated all other programs (five commercial and two non-commercial) at the 1989 London 'computer olympiad' (Hertz et al., 1991, p. 137).

Many of these applications are in their early stages (research is being done on them), and they have not reached commercialisation yet, but they show that a considerable amount of research — and R & D — activity is now being carried out in neural computing. Although it is not yet clear how far the institutionalisation of neural network research will go, and how complementary the neural network and the symbolic paradigms will be, it can now be said that neural computing has emerged as an accepted line of research in its own right. The consensus which resulted from the closure of the perceptron controversy in the late 1960s was broken in the late 1980s, and the neural network controversy reopened. The development of the BP learning algorithm was of great importance for the reopening of the controversy. It was shown earlier in this section that debate about BP it is still going on. At a more general level, there is ongoing debate about the relationships between the symbol

processing approach and the neural network approach (see section 3.4). The future shape of the map of AI and cognitive science is difficult to predict, but one thing seems clear by now: this time round the neural network controversy is not going to end with the rejection of one of the contending positions.

In this section I have looked at Rumelhart and colleagues' back-propagation technique, and I have shown that the neural network controversy reopened after back-propagation was developed in the mid- 1980s. I have discussed the importance of back-propagation, and I have compared Minsky and Papert's renewed criticism of neural networks with their earlier (1969) criticism.

5.3 The neural network explosion

The objective of this section is to characterise briefly some aspects of the growth and institutionalisation of neural network research in the late 1980s. I use the words 'neural network explosion' to refer to the peak in the growth of the neural network research community in the late 1980s. This peak happened approximately between 1986 and 1988. Since then, neural networks has become a research specialty and a research and development (R & D) community in its own right. In this section I characterise the growth of neural network research in terms of conferences, publications, funding, and migration of scientists and engineers to the neural network field. It is important to remember that the growth of neural network research was the consequence of the reopening of the neural network controversy, and not its cause.

The two most serious attempts to quantify the growth of the neural network research community that I am aware of are those carried out by MIT Alfred P. Sloan School of Management researchers Michel Rappa and Koenraad Debackere (1989, 1990). After building an relational database containing 2740 abstracts of journal and conference proceedings papers on neural networks published between 1969 and 1988, Rappa and Debackere (1989, p. 9-10) concluded that the biggest increase in the rate of growth of the neural network research community happened from 1986 to 1988. According to Rappa and his colleague the rate of growth of the community was 60% during those years; the community expanded from 200 members to 1,200 (ibid., p. 9).

In a later study, Rappa and Debackere (1990) carried out a statistical survey of 700 neural network researchers from thirty countries. The survey was done between February and May 1990, and confirmed (ibid., 7) that the first peak in entry of

researchers to the field took place in 1986 and 1987. Of the 700 researchers of the sample, 150 had entered in 1986 and 160 in 1987. (This does not mean that there has been a decline in entry since then. Rappa and Debackere acknowledge [ibid., 7] that the decline they get in their survey after 1987 is explained by the techniques used to identify the sample population.) Rappa and his colleague also concluded that 75% of their respondents had entered the field between 1984 and 1990, whereas only 25% had entered before 1984 (ibid., p. 7).

One of the characteristics of the emergence and institutionalisation of research specialties is the proliferation of scientific conferences. This seems to be a good 'thermometer' of the growth of neural network research too. The Santa Barbara (California) 'Neural Networks for Computing' meeting, organised by the American Institute of Physics (AIP) in 1985, was one of the first neural network meetings, and it showed the interest of the physicists in the field. In section 4.1 it was seen that the migration of physicists to neural networks was a significant feature of the re-emergence of the field in the 1980s. The Santa Barbara meeting had 60 participants (Denker, 1986, preface). One year later, in April 1986, the American Institute of Physics organised its second 'Neural Networks for Computing' conference in Snowbird, Utah (Denker, 1986). 160 people attended the meeting, and the organisers were short of space to accept more participants.

> "We figured that if we made preparations to have two and a half times that many this year, we would be safe. Nevertheless, we were unable . . . to admit all the worthy applicants, which is an indication of the growth of the field." (Denker, 1986, preface)

There is an interesting anecdote about that Snowbird conference which illustrates some features of the emergence of neural networks in the 1980s. Bernard Widrow was 'rediscovered' at the 1986 Snowbird meeting — but he had to raise his hand and introduce himself, because nobody had recognised him. He told me the story when I interviewed him at Stanford University.

"We continued [throughout the 1970s and 1980s] working on this [adaptive filtering and adaptive signal processing] until about 4 years ago, when I heard about a meeting at Snowbird [on neural networks]. So I went to the Snowbird conference. I found people there that was so enthusiastic about the thing [neural networks], there were 200 people there, tremendous support for one another. Instead of hostility and people trying to tear each other down, people were giving papers. I knew that some of them made sense and some of them didn't make sense, but no-one was being terribly critical, everyone was being very supportive. It was like a family, it was like a family reunion, but they were all people whom I had never met before. John Hopfield was there, and I didn't know him at that time, I think that Rumelhart was there although I'm not sure. Terry Sejnowski was definitely there, many many people, key people in the field in the United States, and I think quite a few from overseas, from Europe and Japan, a few from Japan, a number from Europe, definitely. So I knew I had to go back into neural nets. It was funny that someone was giving a paper at the Snowbird meeting, and he said, 'you know, Widrow did that back in 1963,' or something like that. No-one had known me, not a soul, so I thought I should raise my hand, and introduce myself. And so I did, and then everybody wanted to hear about the olden days, so I had lots of people to talk to about the history of neural networks." (Widrow, interview)

Since then Widrow has become a very significant and charismatic member of the neural network research community (he is one of the few researchers who have played an active role in the two peaks of activity in the history of neural networks). In 1988 Widrow was director of the DARPA neural network study (DARPA, 1988), and became president of the International Neural Network Society (INNS).

1987 was the year of the first 'big' neural network conference. After a conference on neural computing in February (Jet Propulsion Laboratories, Pasadena, California) and another one in March (Snowbird, Utah, another one organised by AIP), in June the Institute for Electrical and Electronic Engineers (IEEE) organised

the 'First International Conference on Neural Networks' (ICNN) (Schwartz, 1988, pp. 1 and 11) in San Diego, California. This first ICNN was attended by 1500 people, and there were 25 vendors of neural network technology products. The 'International Neural Network Society' (INNS) was announced at that conference, and by the end of that year it had 1200 members (ibid., p. 11). In the July 1987 annual meeting of the American Association of Artificial Intelligence (AAAI) a workshop on neural networks was organised for the first time, and some technology products related to neural networks were exhibited (ibid., p. 11). There was one more important conference (at least) before the end of 1987: the IEEE 'Conference on Neural Information Processing Systems: Natural and Synthetic,' in Denver, Colorado. That conference was attended by about 750 people (DARPA, 1988, p. xxix). In this 'frantic' context of interest in neural computing, DARPA started its 'Neural Network Study' in October 1987.

Since then a lot of conferences on neural networks have been organised. Rappa and Debackere (1989, p. 28) pointed out that in 1988 and early 1989 about 31 conferences on neural networks were held. One of the most important ones in 1988 was IEEE's second ICNN conference, attended by 2200 people (EE Times, 1988a, p. 49). One of the most exciting moments at that meeting seems to have been Marvin Minsky's talk. Zeitvogel (1988a, pp. 10-11) reported about it in the following terms:

> "Minsky who has been criticized by many for the conclusions he and Papert make in 'Perceptrons,' opened his defense with the line 'Everybody seems to think I'm the devil.' Then he made the statement, 'I was wrong about Dreyfus too, but I haven't admitted it yet,' which brought another round of applause."

These words by Minsky at the ICNN are taken by some people as an 'apology for the damage caused by his 1969 book.' Two things can be said about this. First of all, this talk about apologies is irrelevant from a sociological point of view (and of course it is not clear at all that Minsky has actually said something of the kind). I showed in chapter three that the closure of the

perceptron controversy was a social process, and therefore there is nothing to 'apologise for.' But if 'apologising' means that Minsky has retreated from (or has varied greatly) his position in the controversy, then that is not accurate. According to the report by Zeitvogel (ibid., p. 11) the rest of Minsky's talk went along the lines of his 1988 criticism of back-propagation (and neural networks). This renewed criticism was analysed in section 5.2, and it was then seen that Minsky and Papert are still opposed to many of the (most important) recent claims by neural network researchers. A more important event that occurred at the 1988 ICNN is that DARPA announced its neural network programme (I will come to the funding issue below).

Another important meeting that year was the INNS's first annual conference in September 1988. According to Gail Carpenter, one of the organisers, about 600 papers were presented at the INNS meeting, twice the number of papers presented at the 1988 ICNN (Zeitvogel, 1988b, p. 12). At the INNS meeting Bernard Widrow took the presidency of INNS from Stephen Grossberg. Also in 1988, INNS created the 'Neural Networks' journal, one of the most important publications in neural network research. By 1989, INNS had 3500 members from 38 nationalities (Rappa & Debackere, 1989, p. 27).

The 'DARPA Neural Network Study' (DARPA, 1988), and DARPA's subsequent decision to support (to some degree at least) neural networks helped legitimise neural network research. DARPA's support for neural network research was especially significant because of the agency's role in the development of information processing systems in general — and symbolic AI in particular — in the last decades. According to a report by J. M. Brady, DARPA provided about 75% of funding for AI in the United States in the decade from 1964 to 1974 (Fleck, 1982, pp. 181 and 212). In his report on DARPA's involvement in computer science and engineering in the 1962-1982 period, Arthur Norberg of the Charles Babbage Institute (University of Minnesota) concluded that:

". . . We could point to almost the entire field of artificial intelligence research in the United States as a DARPA affect." (Norberg, 1990, p. 21)

After the 1987 San Diego conference (first ICNN), DARPA decided to carry out a study about neural networks (Schwartz, 1988, p. 11). The study started in October of that year, and it was completed by February 1988. In July 1988 a synopsis of the study was published, and in November the whole study came out (DARPA, 1988). The DARPA neural network study was sponsored by Jasper Lupo, manager of the Automatic Target Recognition project of the Smart Weapons Technology programme (within the Balanced Technology Initiative for the development of 'promising new technologies that could substantially advance US conventional defence capabilities' [DARPA, 1988, p. xxv]). Bernard Widrow was study director, and G. Carpenter, L. Cooper, S. Grossberg, and J. Hopfield were technical consultants.[106] The legitimising role of DARPA's involvement in neural network technology is recognised within the DARPA study itself, in a reference to the agency's lack of support for neural networks in the past.

"Minsky and Papert of MIT pointed out that the perceptron could not solve the 'exclusive-or' class of problems . . . whereupon it and neural network research in general were largely abandoned by DARPA in favor of the apparently more promising realm of symbolic processing." (DARPA, 1988, p. 23)

The DARPA study was finished in February 1988 and, as a result of it, a proposal was made for an eight-year $390 million programme on neural networks (see EE Times, 1988b, p. 1). The

---

106 The final document was written by 10 experts, most of which were working at MIT Lincoln Laboratory (MIT-LL). J. Pearson (SRI) and R. Lippman (MIT-LL) wrote the part on 'Adaptive knowledge processing.' E. Posner (JPL) and T. Goblick (MIT-LL) wrote 'Assessment of neural network technology.' J. Leonard (Hughes Aircraft) and M. Holz (MIT-LL) wrote 'Systems applications.' A. Penz (Texas Instruments) and P. Kolodzy (MIT-LL) wrote 'Simulation/emulation tools and techniques.' D. Psaltis (CalTech) and J. Sage (MIT-LL) wrote 'Advanced implementation technology.'

proposal was revealed at the 1988 ICNN, where Lupo was reported to have said the following:

> "I believe that the technology we are about to embark upon is more important than the atom bomb." (EE Times, 1988b, p. 1)

Lupo's statement reflects the atmosphere of excitement of 1988, but other comments by members of DARPA made clear that there was a very considerable amount of uncertainty about neural networks at the time.

> "DARPA recently announced a major programme on neural net research. It is a gamble, Fields [deputy director of research at DARPA] admits. 'The belief is that even though small neural nets on small problems show small performance, large neural nets on large problems will show large performance,' he says. 'The theory of neural nets is not good enough to support that statement' [he says]. Evidence will therefore be sought through practical experiments." (Durham, 1988, p. 25)

DARPA finally approved a 28-month $33 million 'exploratory seed program' on neural network research (Yoon, 1989, p. 10).[107] Work in projects under that programme was expected to start by July 1989 (Yoon, 1989a, 12-13).[108] One of the three goals of the programme was, significantly, the comparison between neural networks and conventional information processing technologies including symbolic AI, signal processing, and control theory in problems like automatic target recognition and continuous speech recognition (Yoon, 1989,a pp. 10-11; Yoon, interview). The other two goals were the developments of neural network theory and modelling, and neural network hardware implementation technology (Yoon, 1989, p. 10).

---

107 DARPA's 'caution' with neural computing was interpreted by EE Times journalist R. Colin Johnson in the following terms: "The [DARPA] study suggested a $390 million, eight-year research effort . . ., although Widrow had wanted to ask for five times that amount. But after pouring $500 million over the last 10 years into artificial intelligence research, DARPA elected to proceed more cautiously with neural technology" (EE Times, 1988c, p. 26).

108 Later there were some problems because of cuts in the US defence budget (see EE Times 1988c).

By 1989 most of the major US funding agencies had launched programmes in neural networks. Those agencies include NSF (Werbos, interview), ONR (McKenna, interview), AFOSR (Tangney, interview), NASA (Molina, 1990, p. 365) and NIH (ibid.). Neural network programmes were also launched by the European Community ('Basic Research in Adaptive Intelligence and Neural Computing,' BRAIN, starting in 1988-1989, and 'Annie' and 'Pygmalion,' both starting in 1989-1990 within the European Strategic Programme for Research in Information Technology, ESPRIT), and by several European governments (the programme of the West German government, starting in 1988, was the biggest one). Important neural network projects are under way in Japan too in large companies and government laboratories, and it is expected that a neural network programme will be a part of the 'Sixth Generation Computer Programme' to be launched after the Fifth Generation Project comes to an end in 1991 (ibid., p. 366) (for a short review on government funding for neural networks in Japan, United States, and Europe see Johnson & Schwartz, 1990).

Many major information technology companies in the US, Japan, and Europe, as well as small companies specialised in neural networks (mainly in the US), are now developing neural network research and products. In December 1990 it was estimated that there were some 300 vendors of neural network technology products world-wide (ibid., p. 368), although the commercialisation of neural computing is still in its early stages. A detailed study of the latest developments in the commercialisation of neural computing technology products is out of the scope of this dissertation (for a short review, see: Molina, 1990).

The disciplinary distribution (or origins) of researchers is another interesting aspect of the recent growth and institutionalisation of neural network research. In the academic world, there are an increasing number of PhD students in many universities carrying out research on neural networks in departments like AI, cognitive science, electrical engineering,

computer science, and physics. Neural computing is being included as a subject in many postgraduate courses, and research groups and centres dedicated to neural computing are emerging. The current disciplinary distribution of neural network researchers is significantly diverse. In their 1990 study (mentioned earlier in this section) Rappa and Debackere concluded that electrical engineering (34.2%), physical science (19.2%), and computer science (17.8%) were the main disciplines of origin of neural network researchers, with the rest distributed as follows: biological science and engineering 7%, mathematics 6.9%, psychology and cognitive science 5.4%, and neural networks 4.7% (Rappa & Debackere, 1990, p. 12). This distribution corresponds to 1990, and therefore it is not representative of the early years of the emergence of neural networks, let alone of early stages of the history of neural networks. As it was said earlier, most of the respondents in Rappa and Debackere's survey were people who came into the field in the late 1980s ('bandwagoners,' as they put it).

Another feature of the institutionalisation of neural network research is the appearance of scientific journals exclusively dedicated to it. Apart from the earlier mentioned journal of the INNS, called 'Neural Networks,' which came out in 1988, an increasing number of specialised journals dedicated to neural computing have appeared more recently. These include two in 1989: 'Neural Computation' (US, edited by T. Sejnowski, publishes short papers), and 'Connection Science: Journal of Neural Computing, Artificial Intelligence, and Cognitive Research' (Britain, Carfax Publishing Company); and two in 1990: 'Network: Computation in Neural Systems' (Britain, Bristol IOP Publishing), and 'IEEE Transactions on Neural Networks.' There has also been a proliferation of newsletters on neural networks like 'Neural Network Review' and 'Neural Technology Update' (formerly 'Synapse Connection') in the US.

Apart from this, most AI, cognitive science, electronic engineering, and many philosophy journals have published special

issues on neural computing. Examples of this are Cognition (vol. 28, 1988), Brain and Behavioral Sciences (vol. 11, 1988), Southern Journal of Philosophy (vol. 26, suppl., 1987), IEEE Computer (vol. 21, 1988), Journal of Memory and Language (vol. 27, 1988), Artificial Intelligence Review (vol. 3, 1989), Proceedings of the IEEE (vol. 78, 1990), Artificial Intelligence (vol. 46, 1990), AI and Society (vol. 4, 1990). Papers and letters on neural networks appear now regularly in general science journals such as 'Nature' and 'Science' as well as in most 'general' AI, electronic engineering and cognitive science journals. A significant number of textbooks on neural networks have now been published, including Beale and Jackson (1990), Aleksander and Morton (1990), Hecht-Nielsen (1990), and Hertz et al. (1991). A number of books on the foundations of neural networks (like Nadel et al., 1989) and collections of historical papers (e.g. Anderson & Rosenfeld, 1988) have been published. Books on neural networks for the general public have also come out (e.g. Johnson & Brown, 1988; Allman, 1989; Brunak & Lautrup,1990). Recent books on the foundations of AI, like (Graubard, 1988) and (Patridge & Wilks, 1990), dedicate considerable attention to neural network research.

Finally, another interesting aspect of the emergence of neural networks in the 1980s is the 'migration' of researchers of prestige into the field. Some of these researchers have carried out (or are carrying out) important contributions in neural networks. One example of this is the number of Nobel Prize winners in the field: Leon Cooper (Nobel Prize in superconductivity), Francis Crick (Nobel Prize for the 'double helix' of DNA), and Gerald Edelman (Nobel Prize in physiology and medicine).[109] People of great prestige in other fields who are doing leading contributions in neural networks include Carver Mead. Mead's contribution to digital VLSI technology (Mead & Conway, 1980) is widely acknowledged (see for example Feigenbaum & McCorduck, 1984, pp. 51-52). Mead is currently

---

[109] For examples of neural network-related contributions by these researchers see: (Bienenstock, Cooper, & Munro, 1982), (Crick & Asanuma, 1986), (Edelman, 1987).

carrying out leading research in neural networks using analog VLSI (see Mead, 1989). Other people of great prestige within their fields, such as cognitive psychologist George Miller, have emphasised the significance of neural networks.

> "Although some members of the old guard are resistant to these ideas [parallel distributed processing, neural networks], the doyen of the psychology of cognition, Professor George Miller, has remarked that it is the most important revolution in psychology in his day and his day includes the advent of cybernetics, information theory, generative grammar and the digital computer as a tool for simulating the mind." (Sutherland, 1986, p. 486)

In this section I have examined some aspects of the growth of neural network research in the late 1980s. It is important to remember that this growth was the consequence — and not the cause — of the reopening of the neural network controversy in the mid-1980s. This reopening was analysed in earlier sections. I have discussed the growth of neural network research by looking at issues including conferences, publications, number of researchers, migration of researchers to the field, and funding. The growth of neural network research is a consequence of the acceptance of the field. In earlier sections I looked at the critical developments that brought about that acceptance.

5.4 Debate continues

In the preceding sections I showed that developments in the mid-1980s provoked the reopening of the neural network controversy and the revision of the closure of the perceptron controversy. That reopening, and the emergence of neural network research in the late 1980s, brought about a reorganisation of AI and cognitive science research. But although neural network research is now generally accepted as an approach to AI and cognitive science in its own right, there is still ongoing debate as to what the exact 'place' (or interpretation) of neural networks should be. In this section my goal is to show at a general level that this debate about neural network research is still going on and is very much open (I will not analyse this debate in depth). In order to show that, I will discuss three positions (three interpretations of neural networks) that were emerging at the end of the 1980s. I will call them 'implementationism,' 'moderate connectionism,' and 'radical connectionism.'

The 'implementationist' position is the most negative view of neural network research. Researchers in favour of this position claim that neural network research (including the neural network developments of the 1980s) does not suppose any innovation in the explanation of cognition. For them neural computing is at the most a theory about how symbol-processing could be implemented in some kind of non-symbolic substratum. Donald Broadbent (1985) was one of the first researchers who espoused the implementationist interpretation of neural networks. Later, cognitive science researchers Jerry Fodor and Zenon Pylyshyn (1988) made a strong defense of this hypothesis in a paper which created a good deal of controversy.

Broadbent's defense of the implementationist hypothesis was a reaction to a paper by James McClelland and David Rumelhart

(1985) in which these researchers advocated a distributed model of memory. Broadbent situated his argument in the context of David Marr's (1982, ch. 1) notion of 'information processing levels.' Marr distinguished three levels at which an information processing system can be analysed: the level of abstract computational theory, the level of representation and algorithm, and the level of implementation. At the computational level the abstract properties of the information processing task ('mapping from one kind of information to another' [ibid., p. 24]) to be carried out are defined. This includes the goal of the task and the main constituents of the strategy to carry out that goal. At the second level one has to understand the particular type of input and output representation chosen, and the algorithm that is going to be used in order to carry out the input/output transformation. The third level is the level of the physical implementation of the information processing task.

The interaction between Marr's three levels is a very important issue in AI and cognitive science. Marr's late work (Marr, 1982) seems to have been interpreted as implying that each level was independent with respect to the others (see Barrow, 1989, p. 12). I discussed this issue briefly in section 4.1.[110] P. Churchland and T. Sejnowski (1988, p. 741) indicated that Marr's notion of levels had been used to conclude that neuroscience is irrelevant to understanding cognition:

> "Marr (1982) maintained that computational problems of the highest level could be analyzed independently of understanding the algorithm that performs the computation. Similarly, he thought the algorithmic problem of the second level was solvable independently of understanding its physical implementation. Some investigators have used the doctrine of independence to conclude that neuroscience is irrelevant to understanding cognition."

---

[110] It is interesting to note that Marr's earlier work (1969, 1971, 1970) on the modelling of the cerebellum, the hippocampus, and the cerebral neocortex was neural network–like.

Broadbent's (1985) arguments in favour of implementationism are very much on the line of the 'doctrine of independence' of levels of information processing mentioned by Patricia Churchland and Terrence Sejnowski in the quotation above. Broadbent claimed that McClelland and Rumelhart's distributed memory model (i.e. neural network model) was useful only at the level of implementation.

> "McClelland and Rumelhart believe that their approach has implications at the psychological and not merely at the physiological level . . . These claims are not appropriate and might in some circumstances damage the acceptance of the distributed theory at its proper level." (Broadbent, 1985, p. 189)

Broadbent concluded that, because all the computations which can be carried out by a system with distributed representations can also be performed by a system having localist (i.e. discrete, symbolic) representations, the distinction between distributed and localist representations brings nothing new to the study of cognition.

> ". . . Distributed systems are not capable of any computation that cannot also be performed by a system of localized storage . . . The conclusion of the analyses of the 1950s was that the distinction of distributed versus specific representations had no importance at the computational level with which psychology deals." (Broadbent, 1985, p. 190)

Rumelhart and McClelland (1985) replied immediately to Broadbent's criticism. They accused Broadbent of missing the distinction between Marr's representational and computational levels. In Marr's schema — went on these authors — the level of cognitive explanation is the representational-algorithmic level. Therefore, if it is maintained that the neural network approach brings no innovation in the explanation of cognition, it has to be shown that it adds nothing to that level. Important representational level issues, such as how long the computation

takes, and how computation is affected by performance factors do not matter so much at the computational level.

> "At the computational level . . . it does not matter how long the computation takes, or how performance [as opposed to competence] of the computation is affected by performance factors such as memory load, problem complexity, and so on." (Rumelhart & McClelland, 1985, p. 194)

Rumelhart and McClelland indicated that equivalence between two information-processing systems (e.g. a neural network one and a symbolic one) at the abstract computational level is not a very useful criterion at the level of representation and algorithm. They concluded that Broadbent's criticism should not worry neural network researchers too much.

In the atmosphere of emergence of neural network research in the second half of the 1980s, arguments like those of Rumelhart and McClelland against Broadbent's implementationism were gaining strength and momentum. However, implementationist criticism of neural network research did not stop. Two important researchers of the symbol-processing cognitive science community, Jerry Fodor and Zenon Pylyshyn (1988) advocated the implementationist interpretation of neural networks in a paper which created a considerable amount of controversy in the AI-cognitive science community.[111]

Unlike Broadbent, Fodor and Pylyshyn situated their discussion at the level of representation and algorithm, but nonetheless their arguments in favour of the autonomy of the representational level with respect to the implementational level resemble some of the lines of Broadbent's criticism of neural networks. The autonomy of the symbol-processing level with respect to the implementation level is one of the general ideas behind the symbolic paradigm. In the early days of AI and cognitive science, one of the appeals of symbol processing against behaviourism was that it created a level at which one could talk about mental

---

[111] The latest event in that controversy seems to be (Lowever & Rey, 1991).

but still physically realisable (and therefore material) events and processes.

Fodor and Pylyshyn claimed that connectionism is irrelevant at the level of representation and algorithm, and concluded that it does not bring any revolutionary changes to cognitive science.

> " . . . The implementation, and all properties associated with the particular realization of the algorithm that the theorist happens to use in a particular case, is irrelevant to the psychological theory; only the algorithm and the representation on which it operates are intended as psychological hypothesis . . . Given this principled distinction between a model and its implementation, a theorist who is impressed by the virtues of Connectionism has the option of proposing PDP's [neural network systems] as theories of implementation. But then, far from providing a revolutionary new basis for cognitive science, these models are in principle neutral about the nature of cognitive processes." (Fodor & Pylyshyn, 1988, p. 65)

The reasons offered by Fodor and Pylyshyn to support the implementationist hypothesis are based on the priority given by the symbolic paradigm to the logico-syntactic structure of cognitive processes. Their view is that the neural network approach cannot account for one of the basic elements of human cognition: compositionality. Thoughts and mental states have a compositional structure, and cognitive processes depend on that structure. Fodor and Pylyshyn argued that it is not possible to be able (for example) to entertain the thought 'a and b,' and not be able to have the thought 'a,' or to be able to entertain the thought that 'Mary loves Paul' and not be able to entertain the thought that 'Paul loves Mary.' The other side of the property of compositionality is that the same atomic symbol (i.e. 'a') can take part in many symbolic expressions (or composite symbol structures).[112]

---

[112] From the point of view of the sociology of knowledge this seems problematic. The constructivist approach to knowledge characteristic of a good part of social studies of science and technology seems quite far away of the view that 'the same symbol can take part in many different expressions.' The meaning of symbols seems to be rather more contextual and (to a certain extent) socially negotiable than that. But, as I pointed out in

Fodor and Pylyshyn (1988) claimed that neural networks could not explain (or artificially model) compositionality, because the only kind of relationship between the components (units) of a neural network is causal, or numerical, namely the interaction between them through the connecting weights. The kind of reasoning which connectionist models could model would be statistical, which is rather different from the formally or syntactically driven inferential processes favoured within the symbolic approach.[113] But Fodor and Pylyhsyn went on further, and doubted whether association is useful *at all* in studying and modelling cognitive processes.

> ". . . We doubt that much of processing does consist of analyzing statistical relations . . ." (Fodor & Pylyshyn, 1988, p. 68)

For Fodor and Pylyshyn, neural networks would be a theory about the implementation of symbol-processing processes. But how this can be interpreted is unclear to them. They pointed out that trying to implement symbolic processes in massively or fine-grained parallel neural network hardware would cause important problems, and that there are more adequate ways of implementing symbol-processing.

> "We have no principled objection to this view [treating connectionism as an implementation theory] (though there are, as Connectionists are discovering, technical reasons why networks are often an awkward way to implement classical machines). This option would entail rewriting quite a lot of the polemical material in the Connectionist literature, as well as redescribing what the networks are doing as operating on symbol structures, rather than

---

the chapter one, I have chosen not to enter in this kind of debate here. Interestingly, the principle of compositionality has been criticised from cognitive science itself: "We would not want a demonstration that an organism-with-systematicity having encountered 'Lions eat people' as a sentence then knows *ipso facto* that 'People eat lions' is one too. The implausibility of the content often renders people unable to accept (at the simplest crudest level of acceptance) such sentences as sentences . . ." (Wilks, 1990, p. 334).

[113] Contrary to what Fodor and Pylyshyn seem to indicate in their paper, connectionist inferential processes would happen at system level, and not at unit level.

spreading activation among semantically interpreted nodes." (Fodor & Pylyshyn, 1988, pp. 67-68)

It does not seem likely that neural network researchers are going to follow Fodor and Pylyshyn's appeal for a 'rewriting of much of the polemical literature' and for a 'redescription of their systems as operating on symbol structures.' It is important to note that the implementationist interpretation of neural networks is often linked — as in this case — with a 'nothing-has-changed' (in AI and cognitive science) hypothesis. This type of hypothesis is the 'most negative scenario' for neural network researchers.

> "[There] is a real disagreement about the nature of mental processes and mental representations. But it seems to us that it is a matter that was substantially put to rest about thirty years ago; and the arguments that then appeared to militate decisively in favor of the Classical [i.e. symbolic] view appear to us to do so still . . . As far as Connectionist architecture is concerned, there is nothing to prevent minds that are arbitrarily unsystematic. But that result is *preposterous.* Cognitive capacities come in structurally related clusters; their systematicity is pervasive. All the evidence suggests that *punctate minds can't happen.* This argument seemed conclusive against the Connectionism of Hebb, Osgood and Hull twenty or thirty years ago. So far as we can tell, nothing of any importance has happened to change the situation in the meantime." (Fodor & Pylyshyn, 1988, pp. 6 and 49)

Fodor and Pylyshyn are not the only ones who have espoused the 'nothing-has-changed' interpretation of neural network research. Certain remarks by Minsky and Papert (1988, p. vi) in their renewed criticism of neural networks also have some 'nothing-has-changed' flavour.

> "Has not there been a 'connectionist revolution'? . . . Certainly no, in that there has been little clear-cut change in the conceptual basis of the field."

The 'openness' of the debate about neural network research is apparent if one looks at other 'nothing-has-changed' comments. I

include below some comments by Tomaso Poggio, from MIT AI laboratory and Thinking Machines Corporation, a leading machine vision researcher.

> "Poggio . . . jokes about a virus that infects brain scientists, starting a new epidemic every 20 years. The epidemic takes the form of uncritical enthusiasm for a new idea. In the 1920s, the idea was *Gestalt* psychology; in the 1940s, cybernetics; in the 1960s, perceptrons. In the 1980s it is connectionism." (The Economist 1987, p. 94)
>
> " 'Neural networks are accompanied by a lot of irritating hype,' Poggio declares, '. . . Neural nets point out interesting problems, but have not solved the big problems of vision or speech. Ultimately, in my view, when the hype disappears, there's a good possibility they will go the way of perceptrons.' " (Poggio, as quoted by Finkbeiner, 1988, p. 11)

The nothing-has-changed position was defended by leading computer scientists like Daniel Hillis, who developed the 'connection machine' parallel computer. Hillis (1989, pp. 175-176), pointed out that:

> ". . . To build a thinking machine by simply hooking together a sufficiently large network of artificial neurons. The notion of emergence would suggest that such a network, once it reached some critical mass, would spontaneously begin to think. This is a seductive idea because it allows for the possibility of constructing intelligence without first understanding it. Understanding intelligence is difficult and probably a long way off, so the possibility that it might spontaneously emerge from the interactions of a large collections of simple parts has considerable appeal to the would-be builder of thinking machines. Unfortunately, that idea does not suggest a practical approach to construction. The concept of emergence in itself offers neither guidance on how to construct such a system nor insight into why it would work."

The openness of the current debate is quite clear when one looks at the 'nothing-has-changed' views about neural network research. I will not go further in the study of these views here.

My goal was just to show that there is an ongoing debate about the place of neural network research within AI-cognitive science. I will look now at some reactions by neural network researchers against Fodor and Pylyshyn's (1988) implementationist position.

The neural network camp was not long (it could not have been) in replying to Fodor and Pylyshyn's criticism. Paul Smolensky (1987) claimed that, contrary to Fodor and Pylyshyn's argument, connectionism does indeed offer an account of the compositionality of cognitive processes. However, connectionist compositionality is defined in rather different terms. In a connectionist system there are no symbols (in the usual AI sense), but activation patterns. Representations are distributed throughout the parameters of the system. Smolensky argues that connectionist representations, unlike symbolic representations, are sensitive to the context in which they appear (e.g. different activation patterns would correspond to the same word appearing in different contexts). Smolensky claimed that connectionist representations can be decomposed into parts or constituents, but that these simpler parts are not defined in a discrete or symbolic way. They vary in different situations.

One example used by Smolensky (1987) was the expression 'cup with coffee.' This expression would have two parts, namely (the activation pattern corresponding to) 'cup without coffee,' and (the activation pattern corresponding to) 'coffee.' However, in a different situation, the activation patterns corresponding to these 'symbols' would be different.

> "The classical [symbol-processing] and connectionist approaches differ not in whether they accept principles (i) and (ii) [(i) thoughts have composite structure; (ii) mental processes are sensitive to this composite structure], but in how they formally instantiate them . . . In the classical approach, [non-formal] principles (i) and (ii) are formalized using syntactic structures for thoughts and symbol manipulation for mental processes. In the connectionist view (i) or (ii) are formalized using distributed vectorial

> representations for mental states, and the corresponding notion of compositionality, together with association-based mental processes that derive their structure sensitivity from the structure sensitivity of the vectorial representations engaging in those processes." (Smolensky, 1987, p. 151)

Other responses to Fodor and Pylyshyn attacked their 'independence of information-processing levels' (and particularly the independence between the symbol-processing level and the hardware level) assumption. Nick Chater and Mike Oaksford (1990, p. 94) made this clear in their reply to Fodor and Pylyshyn.

> " . . . We . . . consistently urge that the cognitive level must interact with properties of the implementation and so cognitive performance cannot be explained implementation-independently."

Chater and Oaksford speak about connectionist, 'biology-constrained' implementation (and not about brain modelling) here. The degree of neurobiological constraint in neural network architectures varies considerably from system to system, but it seems likely that the interaction between neural network research and neuroscience is going to be one important factor in neural computing in the future.[114]

The second position in the current debate about neural network research that I wanted to discuss in this section could be called 'radical connectionism.' Researchers in favour of this view claim that neural networks will offer an alternative and sufficient way of explaining and modelling most cognitive processes. For them, the symbolic paradigm would only be a useful approximation to the connectionist description — the 'right' description — of cognitive processes. The following comments are examples of this radical connectionist position:

---

[114] This interdisciplinary research area is sometimes called 'computational neuroscience' or 'cognitive neuroscience.' See: (Churchland & Sejnowski, 1988), (Sejnowski, Koch, & Churchland, 1988).

> "It is a mistake to claim that the connectionist approach has nothing new to offer cognitive science. The issue at stake is a central one: Does the complete formal account of cognition lie at the conceptual level? The position taken by the subsymbolic [i.e. neural network] paradigm is: No — it lies at the subconceptual level." (Smolensky, 1988, p. 7)

> ". . . The macroscopic level of description may be only an approximation to the more microscopic theory . . . We view macrolevel theories as approximations to the underlying microstructure that the distributed model presented in our article attempts to capture. As approximations they are often useful, but in some situations it will turn out that a lower level description may bring deeper insight." (Rumelhart & McClelland, 1985, p. 196)

> ". . . We take the symbolic level of analysis to provide us with an approximation to the underlying system. In many cases these approximations will prove useful; in some cases they will be wrong and we will be forced to view the system from the level of units to understand them in detail." (Rumelhart, Smolensky, McClelland, & Hinton, 1986, 56)

PDP researchers emphasised the importance of 'symbolic' concepts such as consciousness, sequential thought, and mental models, but they claimed that these phenomena can be explained with purely connectionist systems (Rumelhart, Smolensky, McClelland, & Hinton, 1986). Rumelhart and his colleagues pointed out that the behaviour of a neural network system can be interpreted in two ways. Each time an input is presented, the system performs a relaxation cycle. At the end of the cycle the activation pattern (energy minimum) most associated with the stimulus is found. This relaxation cycle would take, according to Rumelhart and colleagues, about half a second.[115] Within that time scale, the system works in parallel, but if a larger time scale is considered, the evolution of the system can be seen as a

---

[115] Feldman and Ballard (1982) formulated what the '100 step constraint.' This constraint is an approximate measure of the time required by the human brain to carry out certain complex cognitive processes such as, for example, recognising a human face. I discussed it in section 4.1.

sequence of activation states ('symbols,' but of a connectionist character). Thus Rumelhart and colleagues do not renounce to the study symbolic processes like consciousness or sequential thought, and claim that neural network research will offer a non-symbolic explanation of those processes.[116]

Radical connectionism is a 'programmatic' hypothesis: it poses a research agenda for neural network research for years to come. It remains to be seen to what extent the neural network approach is going to offer useful tools for the study and modelling of the more sequential, structure-sensitive cognitive and intelligent processes. Radical connectionist pronouncements belong to the level of 'bear with us' debating tactics.[117] What is important about the radical connectionist position here is that, in aiming at studying and modelling higher cognitive capabilities, it 'attacks' the symbol-processing approach at its 'core.' The fact that this kind of 'attacks' have been made is a consequence of the confidence and optimism of neural network researchers, and would have been inconceivable a few years ago. For recent developments in neural network studies of 'symbol-processing,' see (Artificial Intelligence, 1990). Of course, the radical connectionist interpretation of neural network research clashes frontally with the implementationist, 'nothing-has-changed'

---

[116] The solution offered by Rumelhart and colleagues to the problem of mental models is somewhat different, although similar in its support for the radical connectionist view. In their account of mental models they propose a second neural network. The first neural network would interact with the world, as it is usually the case with connectionist systems. It would produce outputs when presented with stimuli or inputs. The second neural network, on the other hand, would predict the consequences which the outputs of the first network would have in the world. It would create models of situations of the world, so to speak.These proposals for the study of issues like consciousness, sequential thought, or mental models are rather programmatic, but it is interesting to note that Rumelhart and colleagues do not see the introduction of symbol-processing capabilities in their systems as necessary, even for those 'hard' cases.

[117] Thomas Kuhn (Kuhn, 1970, pp. 157-158) made the following interesting comments about the role of the rhetoric of promise in the early stages of the evolution of a line of research: ". . . The issue [in paradigm debates] is which paradigm should in the future guide research on problems many of which neither competitor can yet claim to resolve completely. A decision between alternate ways of practicing science is called for, and in the circumstances that decision must be based less on past achievement than on future promise. The man who embraces a paradigm at an early stage must often do so in defiance of the evidence provided by problem solving. He must, that is, have faith that the new paradigm will succeed . . ."

view examined earlier in this section. This variety of views on the role of neural network research in AI-cognitive science shows the 'openness' of the current neural network debate.

There is one more position in the neural network debate which I would like to examine in this section. Some researchers take a more eclectic view of the debate and espouse a position that I call 'moderate connectionism.' One of the researchers who has elaborated this position most explicitly is Andy Clark (1989a, 1989b), from the School of Cognitive and Computing Sciences of the University of Sussex (Brighton, England). Clark defended hybrid (partly symbolic, partly connectionist) systems. Clark claimed that (at least) two kinds of theories are needed in order to study and model cognition. On the one hand, for some information-processing tasks (such as pattern recognition) connectionism has advantages over symbolic models. But on the other hand, for other cognitive processes (such as serial, deductive reasoning, and generative symbol manipulation processes) the symbolic paradigm offers in Clark's view adequate models, and not only 'approximations' (contrary to what radical connectionists would claim). Clark claimed that two different types of ontological/methodological frameworks (so to speak) are needed in order to understand (and artificially model) cognition: the symbolic one, and the connectionist one.

> ". . . The computational substrate of human thought comprises (at least) two strands. One, the fast, pattern–seeking operations of a PDP mechanism; the other the slow, serial, gross symbol using, heuristic guided search of classic cognitivism." (Clark, 1989a, p. 63)

> ". . . The kinds of operation we would perform on real, external symbolic structures (and hence the kind we would use in any mental model of the same) are just the operations found in a conventional processor. Operations such as complete copying of a symbol from one location to another, deletion and addition of whole symbols . . . and whole symbol matching operations. In these special cases . . . the conventional model is not any kind of *approximation* to the truth; it is the truth." (Clark, 1989a, pp. 61-62)

Clark conceded that from an evolutionary point of view neural network-like systems 'are earlier.' So neural network-like systems (the brain) had to learn von Neumann-like computation at some point in order to carry out higher level processes. In other words, connectionist systems had to learn to simulate a von Neumann-style architecture for these tasks (mathematical operations like addition or subtraction are a simple example). Clark claimed that, even if neural network-like architectures are more primitive, what matters is not the evolutionary origin of cognitive faculties and processes, but the way those processes 'function' (Clark, 1989a, p. 63; Clark, 1987, p. 13).[118] Because of this, the adequate level of description of certain cognitive capabilities is, in Clark's view, the (implementation-independent) symbolic level, and not the connectionist one (remember his above quoted words: 'it is not an approximation to the truth, it is the truth'). It is not my goal here to discuss Clark's position further, but to describe it as an example of the moderate connectionist position.[119] For my purposes here, it shows (once more) the variety of views about the role of neural network research.

The hybrid approach is also being used from a more practically-oriented perspective by researchers who need both symbol-processing and neural network elements in their systems. An example is Teuvo Kohonen's (1988b) 'neural phonetic typewriter' speech recognition system. The central part of the system is an unsupervised neural network which classifies phonemes. The

---

[118] A parallel (but inverse) example of this would be a conventional (von Neumann) computer simulation of a neural network system. The machine which is carrying out the computation is a serial computer, but the task which is being simulated is better described in neural network terms.

[119] One example is (Estes, 1988). D. Norman (1986, pp. 541 and 543), who played an organisational role in the PDP group, spoke in favour of some kind of symbolic/connectionist hybrid systems: "The PDP system is fine for perception and motor control, fine for categorization. It is possibly exactly the sort of system required for all our automatic, subconscious reasoning. But I think that more is required — either more levels of PDP structures or other kinds of systems — to handle the problems of conscious, deliberate thought, planning, and problem solving . . . [The] weight setting requires some evaluative mechanism that determines when things are going well and when they are not . . . This is one role for deliberate conscious control."

preprocessing part is based on conventional digital signal processing techniques, and the postprocessing part is a symbolic rule base.[120]

In this section I have discussed, from a general point of view, three positions in the current debate about neural networks. These positions were implementationism, radical connectionism, and moderate connectionism. I have shown that the neural network debate is very much open. The neural network controversy reopened in the mid-1980s, and neural network research was accepted as an AI and cognitive science approach in its own right. However, debate about the exact interpretation of neural networks (and the exact place that neural network research should occupy within AI and cognitive science) is still going on, and it is very much open. The definition of the relationships between the symbol-processing and the neural network approaches to AI and cognitive science is still very much a matter open to controversy and negotiation. So far one thing seems clear: in spite of 'nothing-has-changed' claims (see above) the neural network controversy has already reopened, and neural network research has already emerged as an accepted approach to (at least certain problems in) AI and cognitive science. Another thing that seems clear is that the current debate about the place that neural network research should occupy in the study of cognition and in building intelligent machines (the current, in a sense, 'territorial dispute') will shape AI and cognitive science research for years to come.

---

120 Kohonen (1990, p. 1477) warned against insisting too much on distribution: "I am . . . not opposed to the view that neural networks are distributed systems. The massive interconnects that underlie all neural processing are certainly spread over the network; their effects, on the other hand, may be 'focused' on local sites. It seems inevitable, however, that any complex processing task requires *organization of information into separate parts*. Distributed processing models in general underrate this issue. Consequently, many models that process features of input data without structuring exhibit slow convergence and poor generalization ability, usually ensuing from ignorance of the localization of the adaptive processes."

◆ SIX

Conclusion

This 'journey' throughout the history of neural network research is now over. It was a long way, from the times of cybernetics and perceptrons to the revival and the excitement of the late 1980s. When I started my project, the 'terrain' was hilly and rough, difficult to explore. There were no main roads, and there was an infinite number of small paths one could follow. So I had to use my exploratory tools and draw a 'map' with the most important 'places to visit' on it. It is now time to look back to the stages of the journey (the 'places visited'), and discuss the conclusions that can be drawn from each of them and from the journey as a whole. It is also time to reflect on the usefulness of the tools with which the 'map' was drawn, and to discuss other routes which could be followed in the future.

The early period of neural network research, studied in **chapter two** ('Early neural networks') was an exciting time. Many different approaches emerged from the cybernetic concern with the relationships between brain and machine (section 2.1: 'Cybernetics and the origins of neural networks'). Symbol-processing AI and neural networks were two of them, probably the two most powerful, but a great variety of schemes and approaches were developed. It is also interesting to note that the brain/machine problem was originally formulated using neural network terminology (McCulloch & Pitts, 1943). Warren McCulloch and Walter Pitts' aim was to show how logical operations could be supported by a brain-like physical substratum. Neural network terminology was used in different developments related to the brain/machine problem. One example of this is that John von Neumann described the stored-program computer in 1945 using McCulloch and Pitts neurons (Aspray, 1990, p. 173). Another example is that Minsky (1967) wrote a theory of computation with finite machines using McCulloch and Pitts' neural network terminology. In the meantime, neural network researchers started to develop their own approach. Several attempts of combining McCulloch and Pitts neurons with Donald Hebb's notion of learning by synaptic strength

modification were carried out, and the first, exploratory machines were built.

Towards the late 1950s, a more defined neural network approach emerged. The clearest formulation of the driving ideas behind it was done by Frank Rosenblatt (section 2.1: 'The computer and the brain'). Early neural network researchers were explicitly opposed to the use of the von Neumann computer (which, ironically, had been originally described using neural network terminology) as a metaphor for cognition, which is precisely what symbolic AI researchers were doing. Neural network researchers built their own computers, and carried out important tests and experiments. The most important machines of this period were the Mark 1 perceptron, the Madaline, and Minos (section 2.3: 'The perceptron;' section 2.4: 'The Madaline and Minos projects'). At about the same time (late 1950s and early 1960s) the symbolic approach to AI was emerging with increasing momentum (section 2.5: 'The emergence of symbolic artificial intelligence'). Symbolic AI developed in strong association with the digital computer from the very beginning. The digital computer was their simulation and experimentation tool for building intelligent systems. The von Neumann computer showed that one could talk about cognitive entities and processes and still be a 'materialist:' computing programmes and systems were developed that could carry out certain intelligent tasks.

Some conclusions can be drawn from the study of the early period of neural network research. The proliferation of information-processing approaches to the brain/machine problem in the cybernetics movement is amazing. Symbolic AI and neural networks were two of them, but there were many others. More historical and sociological studies of cybernetics and early AI-like research are needed to follow the development of those other approaches. By the late 1950s and early 1960s, symbolic AI and neural networks were the most powerful approaches to the problem of studying cognition and building intelligent machines. Symbolic AI, concentrated in a few centres of excellence and

with privileged access to computational resources, was emerging with increasing momentum. Researchers like McCarthy, Newell, Simon, Minsky and their colleagues and students were developing powerful tools and programs for modelling cognition and building intelligent systems. But neural network research was also being pursued seriously by a significant number of researchers including Rosenblatt's group, Widrow's group, and the group at Stanford Research Institute.

Symbol-processing AI and neural network research were in opposition from the beginning. They were seen by their proponents as two alternative approaches to the problem of studying cognition and building intelligent machines. Furthermore, opposition with researchers favouring the symbol-processing approach has shaped neural network research throughout its history. Susan Star (1989a,pp. 126) pointed out in her study of localisationist brain research that "the shape of localizationist theory was developed through conflict with its opponents." Something similar can be said of neural networks. Conflicts and controversies with researchers favouring symbolic AI have shaped neural network research throughout its history. Authors like Harry Collins (1985), Bruno Latour (1987), and Susan Star (1989a) developed useful elements for the sociological study of scientific controversies. In this dissertation I have tried to combine several of those elements in my study of the history of neural network research.

Chapter three is a study of the first critical moment of controversy of the history of neural network research ('The Perceptron Controversy'). The opposition between the symbol-processing and neural network approaches to AI soon became open controversy. Rosenblatt's perceptron project was funded by the Office of Naval Research and it received considerable attention in the wider society, as well as in the research community. The reaction of researchers favouring symbolic AI was strong. They criticised the perceptron project and its proponents (and neural network research in general) heavily. A

tough and strong controversy about the perceptron (and about the neural network approach as a whole) developed (section 3.1: 'The heat of the controversy'). Controversies are fought with 'rhetorical tactics ' These tactics are used with a view to mobilising and enrolling as many (and as good) allies and resources as possible in one's favour. As Latour (1987) put it, a controversy is a race (a 'proof race'), and the move is not from rhetoric to truth, but from weaker rhetoric to stronger rhetoric. The perceptron controversy was not exception. A great variety of rhetorical tactics were used, and the move was one from weaker to stronger rhetoric.

The debating tactics studied in section 3.1 were only the beginning, the weak rhetoric, so to speak. Stronger rhetoric followed soon. Marvin Minsky and Seymour Papert, from the MIT AI lab, decided to start a project which, if successful, would show the limitations of perceptrons in a decisive manner. 'In a decisive manner' meant that people would stop working on perceptrons once and for all. Minsky and Papert had criticised perceptrons openly before they began their project, and their arguments played an important role in the crisis of early neural networks much before their 1969 book was published.

In the mid-1960s, neural network researchers were having considerable trouble with their machines (section 3.2: 'The crisis of early neural networks'). They were aware that the perceptrons they were studying had important limitations and tried to improve on those aspects. In particular, early neural network researchers were aware that some of the most important limitations of single-layer perceptrons could be overcome with multilayer systems (neural networks with more than one layer of adjustable connections). Researchers tried to develop adequate techniques for modifying the connections of those systems, but they were unsuccessful. Learning techniques comparable in power to those developed for single-layer systems earlier (see section 2.3) were not developed. As controversy increased, some neural network researchers started to change to other projects

of research outside the neural network field. Others, like Rosenblatt and colleagues, continued to work in neural networks, hoping that they could solve some of the limitations of their machines.

When Minsky and Papert's (1969) study was finally published, it was already quite late in the perceptron controversy. However, it is important not to forget that Minsky and Papert's arguments against neural networks were known by the mid-1960s, and that they had been an important factor in the crisis of early neural networks. Minsky and Papert's (1969) 'Perceptrons' book was the result of a project in which they 're-enacted' (Latour, 1987) Rosenblatt's perceptron. 'Perceptrons' contained an elaborated study of the limitations of the single-layer perceptron (a perceptron with one layer of adjustable connections). The interpretative flexibility of this part of Minsky and Papert's study was significant (section 3.3: 'Interpretative flexibility'). The book also contained a pessimistic 'intuitive judgement' about the capabilities of more complex, multilayer perceptrons. The interpretative flexibility of this judgement was great: if an intuitive judgement is not open to interpretative flexibility, then nothing is. Rosenblatt and his colleagues noticed the interpretative flexibility of Minsky and Papert's (1969) arguments, and tried to exploit it in their rhetoric. They also realised that Minsky and Papert's (1969) study was being widely interpreted as showing the uselessness of the neural network approach as a whole, and that closure was approaching (section 3.4: 'Closure').

But let us recapitulate. Which was the situation of the neural network camp when Minsky and Papert's study came out? A quick look at the three main early neural network projects reveals that Rosenblatt did not have many resources and allies. Widrow and colleagues were pursuing engineering applications of their techniques outside neural networks. Rosen and his colleagues at Stanford Research Institute were now working on a robotics project within the symbol-processing approach. The Rosenblatt

camp tried to exploit the interpretative flexibility of Minsky and Papert's study in their favour, but they were increasingly isolated. First, they were isolated in their own field: many AI-oriented neural network researchers, former colleagues, had abandoned neural networks. Secondly, they had failed to enrol the funding agencies, which were key allies in the AI-enterprise. And very importantly: the researchers against neural networks were successful in *linking* their criticism of the perceptron to powerful factors such as the emergence of symbol-processing AI and the development of digital computer technology. This association (or linkage) was a closure mechanism in the perceptron controversy. The Rosenblatt camp was powerless to contest the interpretation of Minsky and Papert's book as showing that neural network research as a whole was not worth pursuing (and therefore that it had to be rejected as an approach to the study and modelling of cognition and to building intelligent machines).

From the point of view of AI-like neural network research, this was defeat for the neural network camp. (Soon the neural network camp lost its most valuable 'ally,'i.e. its 'symbolic leader.' Rosenblatt died in a tragic boating accident in 1971.) But of course the neural network position did not just disappear. Although it retreated from the AI front, neuroscience-oriented research continued, with a relatively strong presence in Europe (see Lighthill, 1973).

Some conclusions can be drawn from chapter three. First, the 'controversy/closure/rhetorical tactics/enrolment of resources and allies' scheme is a useful framework for the interpretation of the developments that shaped neural network research in the 1950s and 1960s. Rosenblatt was increasingly isolated, unable to enrol enough resources and allies to contest the interpretation of Minsky and Papert's (1969) study as showing that perceptrons were not worth pursuing. Another conclusion is that, after closure, the losing side did not disappear. They retreated to other, less spectacular fronts, and continued to develop useful

models: the work carried out throughout the 1970s in content-addressable associative memory and unsupervised neural networks by researchers like C. von der Malsburg, D. Willshaw, T. Kohonen, S. Grossberg, J. Anderson and others is a good example of this. This work was far from the 'hot' centres of AI research in the 1970s, but the situation is quite different today, as those systems are being used and developed further in AI-like research.

After the closure of the perceptron controversy, symbol-processing remained the dominant approach in AI and cognitive science-like research over the years. But in the early 1980s things started to change for neural network research. In **chapter four** I studied some important developments of the early and mid-1980s (Four: 'New Connectionism'). The Parallel Distributed Processing (PDP) researchers did an important job in enrolling resources and allies so as to bring neural network research back to the AI-cognitive science front (section 4.1: Parallel distributed processing'). They linked heterogeneous factors and allies including researchers who had been working in neural networks in the 1970s, researchers who had been working in the symbol-processing approach and had problems in modelling certain intelligent processes, and the trends towards parallel computing of the early 1980s. The transition from symbol-processing AI to neural networks is apparent in some of the early work by the PDP group.

I studied some of the most important innovations of neural network research in the 1980s (i.e. Hopfield's network and the Boltzmann machine) using a 'metaphor scheme' (Barnes, 1974). I showed that such a scheme is useful for the study of the development of those innovations (section 4.2: 'Networks with symmetric connections: metaphors and innovation in neural computing'). The PDP researchers developed further John Hopfield's analogy between neural networks and systems in statistical physics. Concepts like 'energy' and 'temperature' of a neural network were indeed powerful allies. With them, Hinton and colleagues developed a learning technique for multilayer

neural networks, thus giving a solution to one of the most important reverse salients of early neural network research. With allies like 'energy' and 'temperature,' Hinton, Sejnowski, and Ackley were able to reformulate that reverse salient as a (solvable) critical problem (using Thomas Hughes' reverse salient/critical problem terminology). The assumptions they made were rather different from those of earlier neural network researchers like Rosenblatt and Widrow.

The network (in Latour's sense of this term) of new connectionism was quickly getting stronger. And the allies and resources were heterogeneous indeed: 'energy,' 'temperature,' 'stochastic activation function,' as well as those mentioned earlier: researchers who were having trouble in modelling certain problems in the symbol-processing approach, trends towards parallel computing, etc.. Other allies were the System Development Foundation and the Office of Naval Research (ONR). Thomas McKenna (as quoted in Will, 1989, p. 12), from ONR, reminded the forgetful:

> ". . . The PDP group at the University of California, San Diego, were funded at a time when nobody was sure that what they were doing would amount to anything."

Later, after the PDP researchers developed their most important innovations, they were influential in getting the Defense Advanced Research Projects agency (DARPA) — a powerful ally — involved in the neural network enterprise.

But things were happening fast before DARPA got involved in neural networks. In 1986 Rumelhart, Hinton, and Williams of the PDP group developed the back-propagation learning algorithm, a solution to the problem of training multilayer perceptron-like networks. This development was the final catalyser for the reopening of the neural network controversy, as it was seen in **chapter five** (Five: 'Controversy Reopens'). Back-propagation has a long history (section 5.1: 'History of back-propagation'). Paul Werbos had developed a similar algorithm in the 1970s, but found considerable resistance to the *very idea* of applying it to

neural network research. He was isolated, and was not powerful enough to overcome that resistance. Later Rumelhart and his colleagues from the PDP group developed back-propagation as a generalisation of Widrow and Hoff's algorithm for single-layer neural networks, and applied it to multilayer perceptron-like networks successfully (section 5.2: 'Back-propagation: learning in multilayer perceptrons'). Rumelhart and colleagues also found resistance and criticism, but they were able to overcome it. Minsky and Papert realised that the conclusions of their 1969 study — and the closure of the perceptron controversy — were in jeopardy. They counter-attacked, but by then the neural network position was much more powerful than in the late 1960s. Controversy had reopened, and the emergence of neural networks was gaining unstoppable momentum.

Of course, in Star's above mentioned sense, Minsky and Papert's (1988) renewed criticism was a 'positive' contribution to neural network theory. As with their (1969) criticism, Minsky and Papert (1988) helped shape neural network research by indicating where problems lay (and therefore by constructing problems). Nevertheless, the 'nothing has changed' style conclusions contained in Minsky and Papert's (1988) renewed criticism were not widely accepted this time round. The emergence and institutionalisation of neural networks as a research specialty was well under way by then (section 5.3: 'The neural network explosion'). The growth of neural network research in the 1986-1988 period was indeed remarkable, as section 5.3 shows.

As a result of the emergence of neural computing in the late 1980s, the map of AI and cognitive science research is in a process of definition, and there is continuing debate (5.4: 'Debate continues'). The relationships between the symbol-processing and the neural computing approaches are being negotiated and defined, and it is too early to see how the boundaries of the map of AI research are going to be drawn. Certainly the defenders of the 'nothing has changed' position do not seem to be on the

winning side this time round. The recent emergence of connectionism — a demonstration of strength by neural network researchers — goes very much against their claim.

Some conclusions can be drawn from chapter five. One is that, in spite of continuing criticism and controversy, some 'black boxes' have been created in neural network research. The Hopfield network, the Boltzmann machine, and back-propagation are good examples of this. Of course they are continuously being developed in several directions, but in a sense they have already become 'black boxes' or accepted results in neural network research. They are being used as resources — without being criticised — by thousands of researchers. Latour's (1987, pp. 41-42) conclusions apply to those results

> ". . . Few papers are always referred to by later article with similar positive modalities, not only for one generation of articles but for several. This event — extremely rare by all standards — is visible every time a claim made by one article is borrowed without any qualification by many others . . . A black box has been produced . . . A fact is what is collectively stabilised from the midst of controversies when the activity of later papers does not only consist of criticism or deformation but also of confirmation. The strength of the original statement does not lie in itself, but is derived from any of the papers that incorporate it . . . The dissenter will be faced not with one claim in one paper, but with the same claims incorporated in hundreds of papers."

One reason why the history of neural network research is interesting from a sociological point of view is that the construction of black boxes is, as Latour points out, an extremely rare event. But it is also interesting to emphasise that the neural network 'black boxes' — and the whole neural network approach — are still being criticised (and heavily) by researchers in AI. This is even more so because of some aspects of neural network research. One example of this was seen in section 5.2. Rumelhart and colleagues claimed that, although solutions could not be guarantied, back-propagation was working well in all the tests

they had carried out until then. Neural network researchers have been accused of not understanding their systems. It was seen in section 5.4 that, at a more general level, the notion that intelligence emerges from the parallel interaction of many subsymbolic processing units has been (and is being) contested heavily by researchers in AI. Therefore even though some black boxes have been created, debate continues.

But, to a certain extent, neural network researchers can afford to ignore some of those general criticisms this time round, and concentrate on the development of their techniques.[121] At least for the time being, they have room for manoeuvring and 'triangulation' (in Star's [1989a] sense). One frequently used resource in this respect is triangulating between neuroscience-oriented neural networks and information technology-oriented neural networks. Neural network systems are legitimised because of their relevance for brain research, and neuroscience-oriented neural network research is legitimised because of its relevance for computer technology.

Neural network research has just emerged, and many problems remain to be attempted. But, after the emergence, researchers are in a position now to use 'bear with us' tactics in case of trouble. Right now, and for the time being, neural network researchers are in a position to promise 'future results' and be believed. The balance of power in AI-cognitive science has changed significantly from the times of the closure of the perceptron controversy.

Some general conclusions about this sociological study of neural network research can also be drawn. An obvious one is that further research on several issues is needed. In the introduction (chapter one), I described the two main simplifications that I had to do in order to be able to carry out this project. One is that I have concentrated on certain kinds of neural network systems

---

121 Susan Star (1989a, pp. 92-93 and 143-144) described the 'subsuming epistemological questions to debates about technique' and 'ignoring' rhetorical tactics.

(namely supervised neural networks). The other is that I have discussed only the main (and most general) characteristics of symbol-processing AI. It is my claim in this dissertation that this simplification is useful for the interpretation of the main developments of the history of neural network research from the 1950s to the 1980s. But of course future historical projects can complete this study in the the two mentioned directions by looking in more detail at other neural network developments and at the history of symbolic AI.

More historical studies (and of course sociological histories) of AI research are needed, from cybernetics to the present. In a recent review of historical and sociological studies of AI research, Adam (1990) concluded that very little attention has been dedicated to AI from sociology of scientific knowledge approaches. One issue here is that both the sociology of scientific knowledge and AI study knowledge. Because of this, some researchers from the sociology of science tradition have developed direct, 'first level' contributions (so to speak) to AI (e.g. Collins, 1990). This is of course perfectly legitimate, and it is my view that the sociology of knowledge should participate *much more* in the cognitive science enterprise.[122] However, in this project I chose to remain in the traditional 'second level' of the social studies of science and not to study, for example, the relationships between a sociology of knowledge approach to knowledge and a neural network approach to knowledge. That is a different project from the one I have undertaken in this dissertation. Here I have applied a sociological approach to the study of the history neural network research. The fact that a sociological approach can be applied to another approach to the problem of knowledge (neural network research) can only show the strength of the former.

So, what can be learned from this study of the history of neural network research from the point of view of the sociology of

---

[122] Some ideas in this direction are being developed at the Institute for Cognitive Science of the University of California San Diego (Hutchins, interview).

science? I would like to end this chapter by responding to this question. First of all, it can be learned that sociological tools can be applied to the history of neural networks, and that such an application offers a powerful interpretation of the evolution of neural network research. The 'controversy/closure/rhetorical tactics/enrolment of allies and resources' scheme, as developed by authors such as Harry Collins (1985), Bruno Latour (1987), and Susan Star (1989a) offers powerful tools that can be applied to the study of the evolution of neural network research. The priority given by those authors to controversy in the social study of science is, I think, justified. Scientific knowledge is generated and validated through processes of controversy and closure.

The importance of controversy in the history of neural network research is overwhelming. Neural network research has been shaped through controversies with researchers favouring symbol-processing AI. Star (1989a) indicated the importance of debate as a 'positive force' which shapes scientific approaches. The idea is useful at a general level for the interpretation of the evolution of neural network research. It is also especially useful in analysing the problem of training multilayer perceptrons. Minsky and Papert (1969) played a significant role in emphasising the importance of that 'reverse salient' (and thus, in a sense, they helped construct it).

Harry Collins' 'classical' 'controversy/closure' scheme was useful in my study, but had to be completed with other tools. Collins (1985) indicated that rhetorical tactics were always used in the closure of scientific controversies. The idea of closure by using rhetoric was developed further by Bruno Latour (1987). Latour argued that controversies are closed through social processes of rhetoric and power, that is by enlisting heterogeneous allies and actants and by controlling the behaviour of others (a power race). Star's (1989a) idea of science as a heterogeneous and distributed activity whereby local contingencies are eliminated — using diverse tactics — in order

to produce global validity was also useful, and I could find in neural network research instances of many of the tactics she described in her study of localisationist brain research.

Latour's idea of 'from weaker to stronger rhetoric' (by enlisting more allies, resources, and actants) is useful for the study of Minsky and Papert's (1969) move aiming at intervening decisively in the controversy. It was seen that Minsky and Papert's (1969) study can be seen as the result of 're-enacting' (in Latour's sense) Rosenblatt's perceptron. Latour's idea of enrolling and mobilising actants was also useful for the study of the emergence of neural network research in the 1980s, and the PDP group's role in it. Some of the 'associations' made by researchers like Hopfield and the PDP group, like the one between statistical systems and neural networks, were especially important. The metaphor or analogy scheme (Barnes, 1974) is particularly useful for the study of key innovations like the Hopfield network and the Boltzmann machine.

Other conclusions can be drawn from this study which have to do with the controversy/closure/reopening of the controversy pattern of the history of neural network research. The are about the 'social costs' (so to speak) of closing and reopening controversies. One of the premises of the sociology of science, as it was seen in the introduction, is that no knowledge claim or scientific result has an absolute warranty. In other words, in principle, there are always grounds for challenging claims, evidence, or results. Particular instances of 'compelling technical reasons,' 'technical superiority,' and/or 'efficiency' can always, in principle, be challenged (that is, 'universals' do not exist for such things as 'technical superiority'). In other words, the closure of a controversy can always in principle be challenged. But, as in politics, practice is always very different from principle.

There are two examples of this difference between 'in principle' and 'in practice' in the thesis. One is the cost of Minsky and Papert's (1969) move in the early debate. In the early 1960s,

Minsky and Papert sought to make a decisive (i.e. a 'closing') move in the perceptron controversy. The time it took shows that it was a costly process. Another example is the challenge of the closure of the perceptron controversy, and the subsequent reopening of the controversy. The process of enrolment of allies and resources which made that possible *in practice* (that is in spite of the opposition of others) was an extremely costly process in terms of time and 'actants' which had to be mobilised. The time needed was about twenty years. The actants mobilised were many, as it was seen in chapters four and five: the neural network researchers of the 1970s, researchers who were not happy with the symbolic approach to certain problems, especially powerful and innovative associations such as the statistical physics-neural network analogies, other innovations such as back-propagation, powerful digital computers for simulation, developments in parallel computing, physicists, engineers, neuroscientists, the vanguard-like intervention of the PDP group, the funding agencies, etc.. This mobilisation is, I think, a good example of the 'social costs' of reopening a controversy. The 'social costs' of maintaining and developing one's position in a 'proof race' are always increasing. As controversy develops, more allies and actants have to be mobilised, and research becomes costlier and more social.

As section 5.4 shows, controversy continues, as the map of AI and cognitive science is being negotiated. But of course, direct controversy, of the type I have studied in the perceptron controversy and its reopening in the late 1980s cannot go on at those levels forever. In a *direct* confrontation, the costs of maintaining and developing one's position in the controversy are always increasing, as more and more actants are mobilised in order to be able to re-enact your opponent's position and show its flaws and weak points. Such direct controversy is very costly, and cannot be maintained for a long time. Now that both the symbol-processing and the neural network positions are strong enough to resist each other's attacks, they will have to coexist or even co-operate in studying and modelling cognition

and in building intelligent machines, at least for a while. That coexistence and co-operation will happen, as always, through controversies, but they will not be as direct and radical as the ones I have studied in this dissertation. At least for a while.

◆ Appendix 1: Photographs

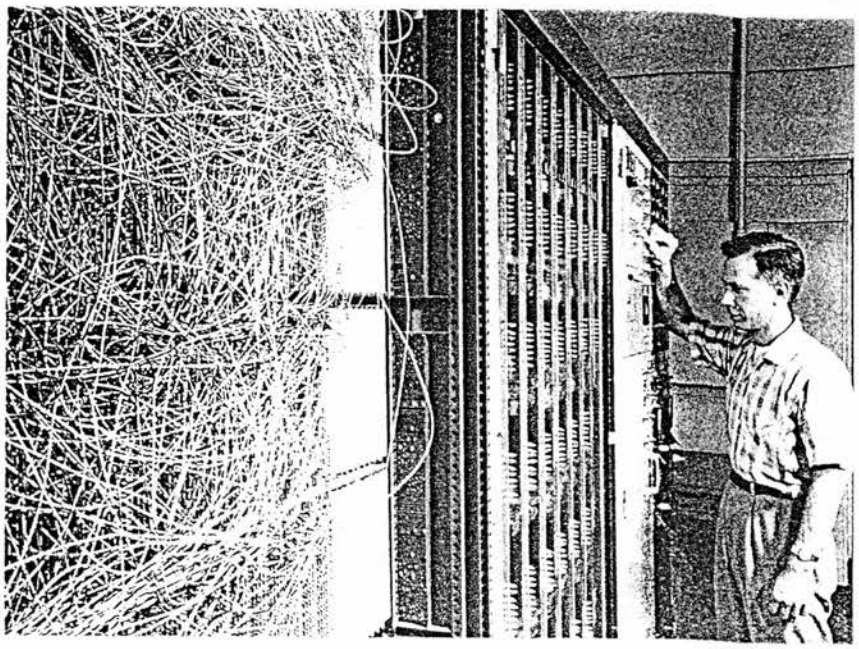
• Photograph 1: Mark 1 Perceptron. On the left, random connections between sensory units and association units. From: (Hecht-Nielsen, 1990, p. 9).

• Photograph 2: Mark 1 Perceptron. Mark 1 Perceptron project engineer Charles Wightman is holding a subrack of 8 motor/potentiometer pairs. Each motor/potentiometer pair functioned as a single adaptive weight. From: (Hecht-Nielsen, 1990, p. 8).

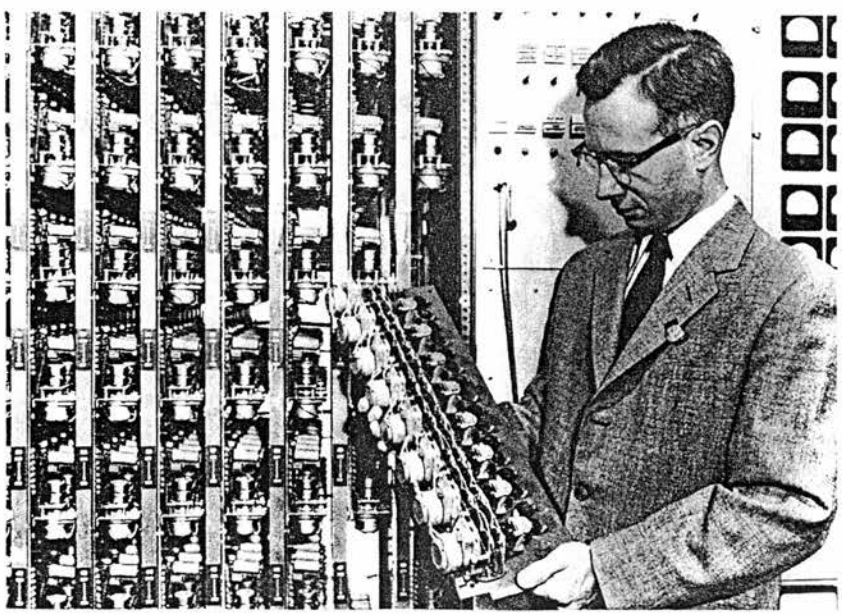
• Photograph 3: Bernard Widrow (Stanford University) holding one of his Adalines.

• Photograph 4: Minos neural network machine, built at Stanford Research Institute. From: (Brain & Munson, 1966, p. 3).

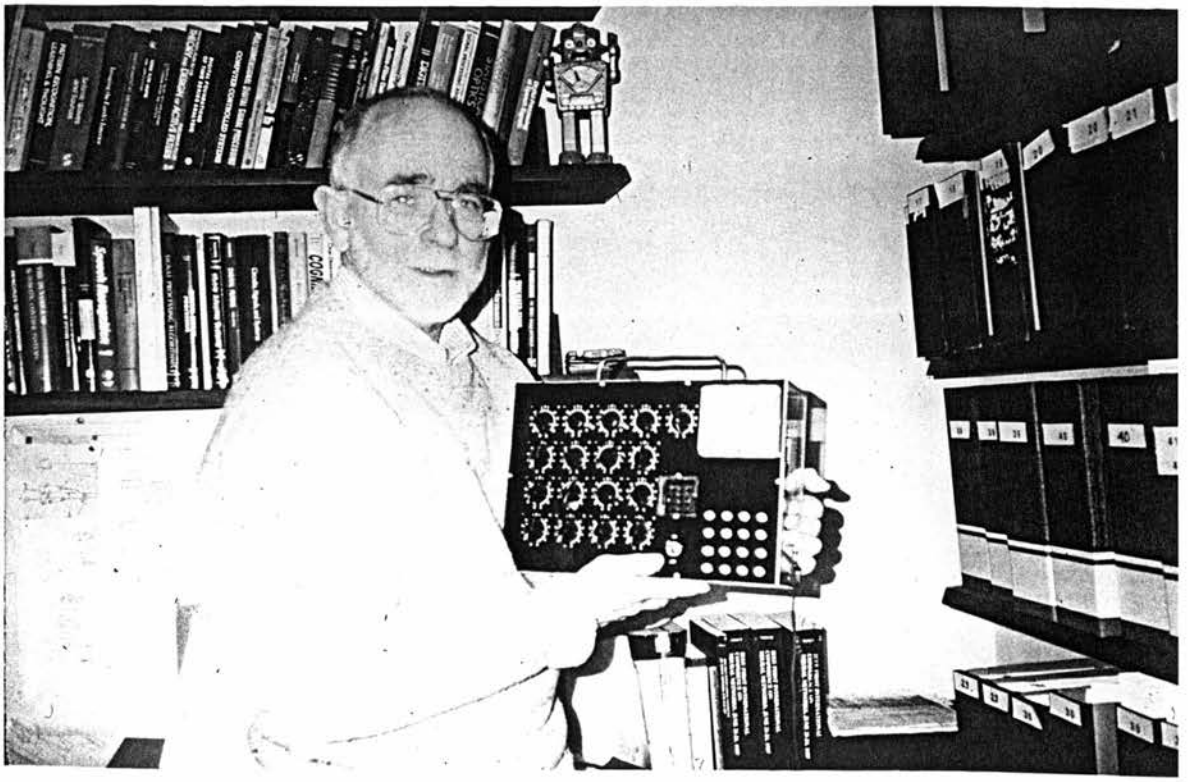
photograph 1



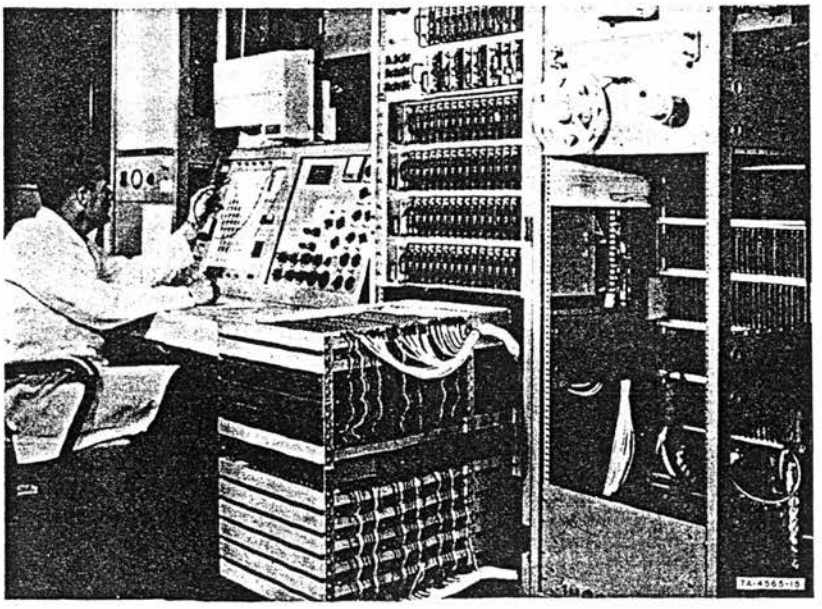
photograph 2



photograph 3



photograph 4



◆ Appendix 2: List of Those Interviewed

Aleksander, Igor. Imperial College of Science, Technology, and Medicine, London. May 15, 1989.

Anderson, James A. Brown University, Providence, Rhode Island. October 20, 1989.

Arbib, Michael A. University of Southern California, Los Angeles, California. November 7, 1989.

Churchland, Patricia S. University of California San Diego, La Jolla, California. November 8, 1989.

Cicourel, A. Cognitive Science, University of California San Diego, La Jolla, California. November 9, 1989.

Denicoff, Marvin. Potomac, Maryland. November 29, 1989 (by telephone).

Duda, Richard. San Jose State University, San Jose, California. November 17 (by telephone).

Feldman, Jerome A. International Computer Science Institute, Berkeley, California. November 10, 1989.

Grossberg, Stephen. Boston University, Boston, Massachusetts. October 18 and 24, 1989.

Hart, Peter. Menlo Park, California. November 19, 1989 (by telephone).

Hopfield, John J. California Institute of Technology, Pasadena, California. November 6, 1989.

Hutchins, E. University of California San Diego, La Jolla, California. November 9, 1989.

Klopf, Harry A. Wright-Patterson Air Force Base, Ohio. November 27, 1989 (by telephone).

Lazzaro, J. California Institute of Technology, Pasadena, California. November 6, 1989.

Licklider, J.C.R. Arlington, Massachusetts. November 30, 1989.

McClelland, James L. Carnegie Mellon University, Pittsburgh, Pennsylvania. November 1, 1989.

McKenna, Thomas. Office of Naval Research, Arlington, Virginia. November 21, 1989.

Mead, Carver A. California Institute of Technology, Pasadena, California. November 6, 1989.

Minsky, Marvin L. Massachusetts Institute of Technology, Cambrigde, Massachusetts. October 25, 1989.

Nilsson, Nils J. Stanford University, Stanford, California. November 3, 1989 (by telephone).

Norman, Donald A. University of California San Diego, La Jolla, California. November 8, 1989.

Papert, Seymour A. Massachusetts Institute of Technology, Cambridge, Massachusetts. December 4, 1989.

Recce, Michael. University College, London. May 16, 1989.

Rosen, Charles. Atherton, California. November 10, 1989.

Rumelhart, David E. Stanford University, Stanford, California. November 13, 1989.

Schwartz, Daniel B. GTE Laboratories, Waltham, Massachusetts. October 26, 1989.

Sejnowski, Terrence J. Salk Institute, San Diego, California. November 8, 1989.

Selfridge, Oliver G. GTE Laboratories, Waltham, Massachusetts. October 27, 1989.

Selviah, Dr. Dept. of Electrical and Electronic Engineering, University College, London. May 16, 1989.

Smolensky, Paul. University of Colorado, Boulder, Colorado. November 14, 1989.

Sutton, Richard S. GTE Laboratories, Waltham, Massachusetts. October 27, 1989.

Tangney, John. Air Force Office for Scientific Research/NL, Washington, DC. November 21, 1989.

Treleavan, Philip. University College, London. May 16, 1989.

von der Malsburg, Christoph. University of Southern California, Los Angeles, California. November 7, 1989.

Werbos, Paul. National Science Foundation, Washington, DC. November 20, 1989.

Widrow, Bernard. Stanford University, Stanford, California. November 13, 1989.

Will, Craig. Institute for Defense Analysis-CSED, Alexandria, Virginia. November 20, 1989.

Williams, Ronald J. Northeastern University, Boston, Massachusetts. November 3, 1989.

Willshaw, David J., MRC, Centre for Cognitive Science, University of Edinburgh, Edinburgh. December 5, 1990.

Yoon, Barbara. Defense Advance Research Projects Agency, DARPA, Arlington, Virginia. November 20, 1989.

Yovits, Marshall. Purdue University, Indianapolis, Indiana. November 28, 1989 (by telephone).

Zipser, David. University of California San Diego, La Jolla, California. November 9, 1989.

◆ Appendix 3: List of Personal Communications by Letter

Gwin, Cecil W. Martins Ferry, Ohio.

O'Brien, Richard D. University of Massachusetts at Amherst.

Rosenblatt, Maurice. Washington DC.

Scattergood, Mark. Englewood, Colorado.

◆ Appendix 4: List of Abbreviations

AFOSR: Air Force Office of Scientific Research
AI: artificial intelligence
AIP: American Institute of Physics
ARPA: see DARPA
ART: adaptive resonance theory
BM: Boltzmann machine
BP: back–propagation
CalTech: California Institute of Technology
CAL: Cornell Aeronautical Laboratory (now Arvin Calspan Advanced Technology Center)
DARPA: Defence Advanced Research Projects Agency
ICNN: International Conference on Neural Networks
IEEE: Institute for Electrical and Electronic Engineers
INNS: International Neural Network Society
IT: information technology`
JPL: Jet Propulsion Laboratory
LMS: least mean square
MIT: Massachusetts Institute of Technology
MIT-LL: Massachusetts Institute of Technology Lincoln Laboratory
NASA: National Aeronautics and Space Administration
NIH: National Institute of Health
NPL: National Physical Laboratory
ONR: Office of Naval Research
PDP: parallel distributed processing
SRI: Stanford Research Institute
VLSI: very large scale integration

◆ References

Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985) A learning algorithm for Boltzmann machines. Cognitive Science, 9 , 147-169 (reprinted in Anderson & Rosenfeld (Ed.),1988, pp. 638-649; quotations in text from reprinted version).

Adam, A. E. (1990) What can the history of AI learn from the history of science? AI and Society, 4 , 232-241.

Aleksander, I., & Morton, H. (1990) An Introduction to Neural Computing. London: Chapman and Hall.

Allman, W. F. (1989) Apprentices of Wonder: Inside the Neural Network Revolution . New York: Bantam Books.

Alternative-Computers. (1989) Time-Life Books: Alternative Computers . Alexandria, Virginia: Time-Life Books.

Anderson, A. (1988) Learning from a computer cat. Nature, 331 , 657-658.

Anderson, J. A. (1972) A simple neural network generating an interactive memory. Mathematical Biosciences, 14 , 197-220.

Anderson, J. A. (interview) Interview. (Brown University, Providence, Rhode Island, 10/10/89).

Anderson, J. A., & Hinton, G. E. (1981) Models of information processing in the brain. In G. E. Hinton, & J. A. Anderson (Ed.), Parallel Models of Associative Memory (pp. 9-48). Hillsdale, New Jersey: Lawrence Erlbaum Associates Inc.

Anderson, J. A., & Rosenfeld, E. (1988) Neurocomputing: Foundations of Research . Cambridge, Massachusetts: The MIT Press.

Apter, M. J. (1972) Cybernetics: a case study of a scientific subject-complex. The Sociological Review Monograph, 18 , 93-115.

Arbib, M. A. (1983) Cognitive science: the view from brain theory. In F. Machlup, & U. Mansfield (Ed.), The Study of Information: Interdisciplinary Messages (pp. 81-91). New York: John Wiley & Sons, Inc.

Arbib, M. A. (1987) Brains, Machines, and Mathematics (second ed.). New York: Springer-Verlag.

Arbib, M. A. (1989) The Metaphorical Brain 2: Neural Networks and Beyond . New York: John Wiley & Sons.

Arbib, M. A. (1989) Schemas and neural networks for sixth generation computing. Journal of Parallel and Distributed Computing, 6 , 185-216.

Arbib, M. A. (interview) Interview. (University of Southern California, Los Angeles, 7/11/89).
Artificial-Intelligence. (1990) Special issue on connectionist symbol processing. Artificial Intelligence, 46 , issues 1-2.
Ashby, W. R. (1952) Design for a Brain . New York: Wiley.
Aspray, W. (1990) John von Neumann and the Origins of Modern Computing . Cambridge, Massachusetts: The MIT Press.
Ballard, D. H., Hinton, G. E., & Sejnowski, T. J. (1983) Parallel visual computation. Nature, 306 , 21-26.
Barber-Associates-Inc. (1975). The Advanced Research Projects Agency 1958-1974 . Richard J. Barber Associates, Inc.; 1000 Connecticut Av.; Washington DC 20036.
Barnes, B. (1974) Scientific Knowledge and Sociological Theory . London: Routledge & Kegan Paul.
Barnes, B. (1977) Interests and the Growth of Knowledge . London: Routledge & Kegan Paul Ltd.
Barnes, B. (1982) T. S. Kuhn and Social Science . London: The Macmillan Press Ltd.
Barnes, B. (1983) On the conventional character of knowledge and cognition. In K. Knorr-Cetina, & M. Mulkay (Ed.), Science Observed: Perspectives on the Social Study of Science (pp. 19-51). London: Sage Publications.
Barnes, B., & Edge, D. (1982) Science in Context: Readings in the Sociology of Science. Milton Keynes, England: The Open University Press.
Barrow, H. G. (1989) AI, neural networks, and early vision. AISB Quarterly, 69 , 6-25.
Beale, R., & Jackson, T. (1990) Neural Computing: An Introduction . Bristol, England: Adam Hilger.
Bernstein, J. (1981) Profiles: AI, Marvin Minsky. The New Yorker, December 14, 50-126.
Bienenstock, E. L., Cooper, L. N., & Munro, P. W. (1982) Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. Journal of Neuroscience, 2 , 32-48.
Bijker, W. E., Hughes, T. P., & Pinch, T. (1987) The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology. Cambridge, Massachusetts: The MIT Press.
Block, H. D. (1962) The perceptron: a model for brain functioning, I. Reviews of Modern Physics, 34 , 123-135 (Reprinted in J. A. Anderson & E. Rosenfeld (Ed.), 1988 Neurocomputing: Foundations of Research (pp. 138-150);

Cambridge, Massachusetts: The MIT Press; quotations in text from reprinted version).

Block, H. D. (1970) A review of 'Perceptrons'. Information and Control, 17 , 510-522.

Block, H. D., Knight, B. W., & Rosenblatt, F. (1962) Analysis of a four layer series-coupled perceptron. Review of Modern Physics, 34 , 135-142.

Bloor, D. (1976) Knowledge and Social Imagery. London: Routledge & Kegan Paul Ltd.

Bobrow, D. G., & Norman, D. A. (1975) Some principles of memory schemata. In D. G. Bobrow, & A. Collins (Ed.), Representation and Understanding: Studies in Cognitive Science (pp. 131-149). New York: Academic Press.

Brain, A., Forsen, G., Hall, D., & Rosen, C. (1963). A large, self-contained learning machine. A paper presented at the 1963 Western Electronic Show and Convention, San Francisco, California, August 20-23 (exhibit). Stanford Research Institute, Stanford, California.

Brain, A. E. (1961) The simulation of neural elements by electrical networks based on multi-aperture magnetic cores. Proceedings of the IRE, 49 January, 49–52.

Brain, A. E., & Munson, J. H. (1966). Graphical-data-processing research study and experimental investigation, prepared for the US Army Electronics Command (Fort Monmouth, New Jersey) (contract DA 36–039 AMC–03247, final report). Stanford Research Institute, Menlo Park, California.

Broadbent, D. (1985) A question of levels: comment on McClelland and Rumelhart. Journal of Experimental Psychology: General, 114 2, 189–192.

Brown, R. J. (1964) Adaptive multiple-output threshold systems and their storage capacities . Thesis, tech. rep. 6771-1, Stanford Electron. Labs., Stanford, California.

Brunak, S., & Lautrup, B. (1990) Neural Networks: Computers with Intuition . Singapore: World Scientific.

Bryson, A. E., & Ho, Y.-C. (1969/1975) Applied Optimal Control . New York: Hemisphere Publishing, 1975 (revised printing if the 1969 edition).

Carpenter, G. A. (1989) Neural network models for pattern recognition and associative memory. Neural Networks, 2 , 243-257.

Carpenter, G. A., & Grossberg, S. (1987) A massively parallel architecture for a self-organizing neural pattern recognition machine. Computer Vision, Graphics, and Image Processing, 37 , 54-115.

Carpenter, G. A., & Grossberg, S. (1988) The ART of adaptive pattern recognition by a self-organizing neural network. IEEE Computer, March 1988, 77-88.

Chater, N., & Oaksford, M. (1990) Autonomy, implementation and cognitive architecture: a reply to Fodor and Pylyshyn. Cognition, 34, 93-107.

Churchland, P. S., & Sejnowski, T. J. (1988) Perspectives on cognitive neuroscience. Science, 242, 741-745.

Clark, A. (1987) Connectionism and cognitive science. In J. Hallam, & C. Mellish (Ed.), Advances in Artificial Intelligence. Proceedings of the 1987 AISB Conference, University of Edinburgh, 6-10 April (pp. 3-15). Chichester, Great Britain: John Wiley and Sons.

Clark, A. (1989a) Connectionism and the multiplicity of mind. Artificial Intelligence Review, 3, 49–65.

Clark, A. (1989b) Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing. Cambridge, Massachusetts: The MIT Press.

Cognitive-Science. (1985) Special issue on connectionist models and their application. Cognitive Science, 9, 1-169.

Collins, H. M. (1975) The seven sexes: a study in the sociology of a phenomenon, or the replication of experiments in physics. Sociology (Journal of British Sociological Association), 9, 205-224 (reprinted in Barnes & Edge (Ed.) 1982, pp. 94-116; quotations in main text from reprinted version).

Collins, H. M. (1981a) Stages in the Empirical Programme of Relativism. Social Studies of Science, 11, 3-11.

Collins, H. M. (1981b) Son of seven sexes: the social destruction of a physical phenomenon. Social Studies of Science, 11 1, 33-62.

Collins, H. M. (1981c) Knowledge and Controversy: Studies of Modern Natural Science . London: Sage Publications (Special issue of Social Studies of Science, vol. 11. no. 1, February 1981).

Collins, H. M. (1983) An empirical relativist programme in the sociology of scientific knowledge. In K. D. Knorr–Cetina, & M. Mulkay (Ed.), Science Observed (pp. 85-113). London: SAGE Publications Ltd.

Collins, H. M. (1985) Changing Order: Replication and Induction in Scientific Practice . London: SAGE Publications Ltd.

Collins, H. M. (1990) Artificial Experts: Social Knowledge and Intelligent Machines . Cambridge, Massachusetts: The MIT Press.

Congressional-Record. (1971) Tribute to Dr Frank Rosenblatt . United States Congressional Record: Proceedings and Debates of the 92d Congress, July 28 1971.

Cover, T. M. (1964) Geometrical and statistical properties of linear threshold devices . PhD thesis, Tech. rep. 6107-1, Stanford Electron. Labs., Stanford, California.

Cowan, J. D., & Sharp, D. H. (1987). Neural nets (LA-UR-87-4098). Los Alamos National Laboratory, Los Alamos, New Mexico.

Cowan, J. D., & Sharp, D. H. (1988) Neural nets and artificial intelligence. In S. R. Graubard (Ed.), The Artificial Intelligence Debate: False Starts, Real Foundations (pp. 85-121). Cambridge, Massachusetts: The MIT Press.

Craik, K. J. W. (1943) The Nature of Explanation . Cambridge, England: Cambridge University Press.

Crick, F. H. C., & Asanuma, C. (1986) Certain aspects of the anatomy and physiology of the cerebral cortex. In J. L. McClelland, D. E. Rumelhart, & The-PDP-Research-Group (Ed.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 2, Psychological and Biological Models (pp. 333-371). Cambridge, Massachusetts: The MIT Press.

Cruz, C. A. (1988) Understanding Neural Networks: A Primer . Amherst, New Hampshire: Graeme Publishing Corporation.

DARPA. (1988) Darpa Neural Network Study . Fairfax, Virginia: Armed Forces Communications and Electronics Association (AFCEA) International Press.

Darrach, B. (1970) Meet Shaky, the first electronic person. Life, 69 21, November 20, 58b-68.

de Mey, M. (1982) The Cognitive Paradigm . Dordrecht, Holland: D. Reidel Publishing Company.

Denicoff, M. (interview) Interview . (Potomac, Maryland, 29/11/89, phone interview).

Denker, J. S. (1986) Neural Networks for Computing . New York: American Institute of Physics.

Deutsch, K. W. (1963) The Nerves of Government: Models of Political Communication and Control . New York: Free Press.

Dickson, D. (1988) The New Politics of Science . Chicago: The University of Chicago Press.

Duda, R. O. (interview) Interview . (San Jose State University, San Jose, California, 17/11/89, phone interview).

Duda, R. O., & Hart, P. E. (1973) Pattern Classification and Scene Analysis . New York: John Wiley and Sons.

Durbin, R., & Willshaw, D. J. (1987) An analogue approach to the travelling salesman problem using an elastic net method. Nature, 326 , 689-691.

Durham, T. (1988) A billion dollar brain with a one-track mind? Computing, 8 December, pp. 24-25.

Edelman, G. M. (1987) Neural Darwinism . Oxford, England: Oxford University Press.

Edge, D. O., & Mulkay, M. J. (1976) Astronomy Transformed: The Emergence of Radio Astronomy in Britain . New York: John Wiley & Sons.

EE-Times. (1988a) Smart thinking shows at neural conference. Electronic Engineering Times, August 15, pp. 49, 65, and 68.

EE-Times. (1988b) DARPA backs neural nets. Electronic Engineering Times, August 8, pp. 1 and 96.

EE-Times. (1988c) DARPA neural network budget imperiled. Electronic Engineering Times, September 4, p. 26.

Estes, W. K. (1988) Toward a framework for combining connectionist and symbol-processing systems. Journal of Memory and Language, 27 , 196-212.

Fahlman, S. E. (1981) Representing implicit knowledge. In G. E. Hinton, & J. A. Anderson (Ed.), Parallel Models of Associative Memory (pp. 145-160). Hillsdale, New Jersey: Lawrence Erlbaum Associates Inc.

Fahlman, S. E., & Hinton, G. E. (1987) Connectionist architectures for artificial intelligence. IEEE Computer, January , 100-109.

Farley, B. G., & Clark, W. A. (1954) Simulation of self-organizing systems by digital computer. IRE Transactions on Information Theory, 4 , 76-84.

Feigenbaum, E. A., & Feldman, J. (1963) Computers and Thought . New York: McGraw-Hill Book Company.

Feigenbaum, E. A., & McCorduck, P. (1984) The Fifth Generation: Artificial Intelligence and Japan's Computer Challenge to the World (revised and updated ed.). New York: The New American Library of Canada Ltd.

Feldman, J. A. (1981) A connectionist model of visual memory. In G. E. Hinton, & J. A. Anderson (Ed.), Parallel Models of Associative Memory (pp. 49-82). Hillsdale, New Jersey: Lawrence Erlbaum Associates Inc.

Feldman, J. A. (1988) Connectionist representation of concepts. In D. Waltz, & J. A. Feldman (Ed.), Connectionist Models and Their Implications (pp. 341-363). Norwood, NJ: Ablex Publishing Corporation.

Feldman, J. A. (interview) Interview. (International Computer Science Institute, Berkeley, California, 10/11/89),

Feldman, J. A., & Ballard, D. H. (1982) Connectionist models and their properties. Cognitive Science, 6 , 205-254 (reprinted in Anderson and Rosenfeld, (Ed.),1988, pp. 484-507, quotations in text from reprinted version).

Feldman, J. A., Fanty, M. A., & Goddard, N. H. (1988) Computing with structured neural networks. IEEE Computer, March , 91-103.

Finkbeiner, A. (1988) The brain as template. Mosaic, 19 2, 3-15.

Fleck, J. (1978) The structure and development of artificial intelligence: a case study in the sociology of science . Unpublished MSc dissertation, University of Manchester.

Fleck, J. (1982) Development and establishment in artificial intelligence. In N. Elias, H. Martins, & R. Whitley (Ed.), Scientific Establishments and Hierarchies. Sociology of the Sciences, vol. VI (pp. 169-217). Dordrecht, Holland: D. Reidel Publishing Company.

Fleck, J. (1984) Artificial intelligence and industrial robots: An automatic end for utopian thought. In E. Mendelsohn, & H. Nowotny (Ed.), Nineteen Eighty-Four: Science between Utopia and Dystopia: Sociology of the Sciences, vol. viii (pp. 189-231). Dordretch, Holland: Reidel Publishing Company.

Fleck, J. (1987) Postscript: The commercialisation of artificial intelligence. In B. P. Bloomfield (Ed.), The Question of AI (pp. 149-164). London: Croan-Helm.

Fodor, J. A., & Pylyshyn, Z. W. (1988) Connectionism and cognitive architecture: a critical analysis. Cognition, 28 , 3-71.

Fodor, J. E. (1975) The Language of Thought . New York: Thomas Y. Crowell.

Forsyth, R. (1989) The brain mimics are back in business. The Guardian, 12 January, p. 25.

Gardner, H. (1985) The Mind's New Science: A History of the Cognitive Revolution . New York: Basic Books, Inc.

Geman, S., & Geman, D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 6 , 721-741.

Graham, G. (1987) Connectionism in Pavlovian Harness. The Southern Journal of Philosophy, 26 supplement, 73-91.

Grossberg, S. (1967) Nonlinear difference-differencial equations in prediction and learning theory. Proc. Nat. Acad. Sci. USA, 58 , 1329-1334.

Grossberg, S. (1968) Some nonlinear networks capable of learning a spacial pattern of arbitrary complexity. Proc. Nat. Acad. USA, 59 , 368-372.

Grossberg, S. (1969a) Some networks that can learn, remember, and reproduce any number of complicated space-time patterns, Part I. J. Math. Mech., 19 , 53-91.

Grossberg, S. (1969b) On the serial learning of lists. Math. Biosci., 4 , 201-253.

Grossberg, S. (1969c) On the production and release of chemical transmitters and related topics in cellular control. J. Theor. Biol., 22 , 325-364.

Grossberg, S. (1970) Some networks that can learn, remember, and reproduce any number of complicated space-time patterns, part II. Stud. Appl. Math., 49 , 135-166.

Grossberg, S. (1976a) On the development of feature detectors in the visual cortex with applications to learning and reaction-diffusion systems. Biological Cybernetics, 21 , 145-159.

Grossberg, S. (1976b) Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. Biological Cybernetics, 23 , 121-134.

Grossberg, S. (1976c) Adaptive pattern classification and universal recoding II: feedback, expectation, olfaction, and illusions. Biological Cybernetics, 23 , 187-202.

Grossberg, S. (1978) A theory of visual coding, memory, and development. In E. L. J. Leeuwenberg, & H. F. J. M. Buffart (Ed.), Formal Theories of Visual Perception New York: Wiley.

Grossberg, S. (1987) Competitive learning: from interactive activation to adaptive resonance. Cognitive Science, 11 , 23-63.

Grossberg, S. (1989) Perceptrons (book review). AI Magazine, summer, 91-92.

Grossberg, S. (interview) Interview. (Boston University, Massachusetts,18 and 24/10/89),

Minsky, M. L. (1968) Introduction. In M. L. Minsky (Ed.), Semantic Information Processing (pp. 1-32). Cambridge, Massachusetts: The MIT Press.

Hagstrom, W. O. (1965) The Scientific Community . Carbondale, Illinois: Southern Illinois University Press.

Harmon, L. D., & Lewis, E. R. (1966) Neural modeling. Physiological Reviews, 46 , 513-591.

Hart, P. E. (interview) Interview . (Menlo Park, California, 29/11/89, phone interview).

Harvey, B. (1981) Plausibility and evaluation if knowledge: a case study of experimental quantum mechanics. Social Studies of Science, 11, 95-130.

Hawkins, J. K. (1961) Self-Organizing Systems: A Review and Commentary. Proceedings of the Institute of Radio Engineers (IRE), 49 January, 31-48.

Hay, J. C. (1960). Mark I perceptron operators manual (VG-1196-G-5). Cornell Aeronautical Laboratory, Buffalo, New York.

Hay, J. C., Martin, F. C., & Wightman, C. W. (1960) Record of IRE 1960 National Convention, part 2 , New York.

Hebb, D. O. (1949) The Organization of Behavior . New York: Wiley.

Hecht-Nielsen, R. (1990) Neurocomputing . Reading, Massachusetts: Addison-Wesley.

Hertz, J., Krogh, A., & Palmer, R. G. (1991) Introduction to the Theory of Neural Computation . Redwood City, California: Addison-Wesley Publishing Company.

Hesse, M. (1963) Models and Analogies in Science . London: Sheed & Ward.

Hillis, W. D. (1989) Intelligence as emergent behavior; or the songs of Eden. In S. R. Graubard (Ed.), The Artificial Intelligence Debate: False Starts, Real Foundations (pp. 175-189). Cambridge, Massachusetts: The MIT Press.

Hinton, G. E. (1977) Relaxation and its role in vision . Unpublished PhD dissertation, University of Edinburgh, Scotland.

Hinton, G. E. (1981) Implementing semantic networks in parallel hardware. In G. E. Hinton, & J. A. Anderson (Ed.), Parallel Models of Associative Memory (pp. 161-188). Hillsdale, New Jersey: Lawrence Erlbaum Associates Inc.

Hinton, G. E., & Anderson, J. A. (1981) Parallel Models of Associative Memory . Hillsdale, New Jersey: Laurence Erlbaum Associates Inc.

Hinton, G. E., & Anderson, J. A. (1989a) Parallel Models of Associative Memory (updated ed.). Hillsdale, New Jersey: Laurence Erlbaum Associates Inc.

Hinton, G. E., & Anderson, J. A. (1989b) Introduction to updated edition. In G. E. Hinton, & J. A. Anderson (Ed.), Parallel Models of Associative Memory (pp. 1-13). Hillsdale, New Jersey: Lawrence Erlbaum Associates Inc.

Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986) Distributed representations. In D. E. Rumelhart, J. L. McClelland, & The-PDP-Research-Group (Ed.), Parallel Distributed Processing: Explorations in the Microstructure of

Cognition, vol. 1. Foundations (pp. 77-109). Cambridge, Massachusetts: The MIT Press.

Hinton, G. E., & Sejnowski, T. J. (1986) Learning and relearning in Boltzmann machines. In D. E. Rumelhart, J. L. McClelland, & The-PDP-Research-Group (Ed.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 1. Foundations (pp. 282-317). Cambridge, Massachusetts: The MIT Press.

Hinton, G. E., Sejnowski, T. J., & Ackley, D. H. (1984). A learning algorithm for Boltzmann machines (CMU-CS-84-119). Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Pennsylvania.

Hockney, R. W., & Jesshope, C. R. (1988) Parallel Computers 2: Architecture, Programming, and Algorithms . Bristol, England: Adam Hilger.

Hoff, M. E. (1962) Learning phenomena in networks of adaptive switching circuits . PhD thesis, tech. rep. 1554-1 Stanford Electron. Labs., Stanford, California.

Hopfield, J. J. (1982) Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences, 79 , 2554-2558 (reprinted in Anderson and Rosenfeld (Ed.) 1988, pp. 460-464; quotations in text from reprinted version).

Hopfield, J. J. (1984) Neurons with graded response have collective computational properties like those of two-state neurons. Proceedings of the National Academy of Sciences, 81 , 3088-3092.

Hopfield, J. J. (interview) Interview . (California Institute of Technology, Divisions of Chemistry and Biology, Pasadena, California, 6/11/89).

Hopfield, J. J., & Tank, D. W. (1986) Computing with neural circuits: a model. Science, 233 , 625-633.

Hubel, D. H., & Wiesel, T. N. (1959) Receptive fields of single neurones in the cat's striate cortex. Journal of Physiology, 148 , 574-591.

Huber, W. A. (1967). Learning machine techniques for pattern classification (typescript). US Army Electronics Command, Fort Monmouth, New Jersey.

Hughes, T. P. (1983) Networks of Power: Electrification in Western Society, 1880-1930 . Baltimore, Maryland: Johns Hopkins University Press.

Hummel, R. A., & Zucker, S. W. (1983) On the foundations of relaxation labelling processes. IEEE Transactions on Pattern Analysis and machine Intelligence, 5 , 267-287.

Hutchins, E. (interview) Interview. University of California San Diego, Cognitive Science, 9/11/89.

Jeffress, L. A. (1951) Cerebral Mechanisms in Behavior. The Hixon Symposium . New York: John Wiley & Sons, Inc.

Johnson, R. C., & Brown, C. (1988) Cognizers: Neural Networks and Machines that Think . New York: John Wiley and Sons, Inc.

Johnson, R. C., & Schwartz, T. J. (1990) IJCNN gov't panel: Governments fund neural nets worldwide. Neural Technology Update (formerly Synapse Connection), 4 2, pp. 1 and 5-7.

Johnson-Laird, P. N. (1988) The Computer and the Mind: An Introduction to Cognitive Science . London: Fontana Press.

Kabrisky, M. (1966) A Proposed Model for Visual Information Processing in the Human Brain . Urbana, Illinois: University of Illinois Press.

Kanerva, P. (1988) Sparse Distributed Memory . Cambridge, Massachusetts: The MIT Press.

Kirkpatrick, S., Gelatt, C. D. J., & Vecci, M. P. (1983) Optimization by simulated annealing. Science, 220 , 671-680.

Knorr-Cetina, K. D., & Mulkay, M. (1983) Science Observed: Perspectives on the Social Study of Science . London: Sage publications Ltd.

Kohonen, T. (1977) Associative Memory — A System Theoretic Approach . Berlin: Springer-Verlag.

Kohonen, T. (1982) Self–organized formation of topologically correct feature maps. Biological Cybernetics, 43 , 59-69.

Kohonen, T. (1984) Self-organization and Associative Memory . Berlin: Springer-Verlag.

Kohonen, T. (1988a) Self-Organization and Associative Memory (second ed.). New York: Springer-Verlag.

Kohonen, T. (1988b) The "neural" phonetic typewriter. IEEE Computer, March 1988, 11-22.

Kohonen, T. (1990) The self-organizing map. Proceedings of the IEEE, 78 9, 1464-1480.

Konheim, A. G. (1963) A geometric convergence theorem for the perceptron. J. Soc. Indust. Appl. Math., 11 , 1-14.

Kuhn, T. S. (1970) The Structure of Scientific Revolutions (second, enlarged ed.). Chicago, Illinois: The University of Chicago Press.

Kuhn, T. S. (1977) The Essential Tension: Selected Studies in Scientific Tradition and Change . Chicago, Illinois: University of Chicago Press.

Larson, E. (1986) Neural chips. Omni, November, 113-169.

Lashley, K. S. (1929) Brain Mechanisms and Intelligence . Chicago, Illinois: University of Chicago Press.

Lashley, K. S. (1950) In search of the engram. Society of Experimental Biology Symposium, No. 4, Psychological Mechanisms in Animal Behavior (pp. 454-455, 468-473, and 477-480). Cambridge, England: Cambridge University Press. (Reprinted in J. A. Anderson & E. Rosenfeld (Ed.), 1988 Neurocomputing: Foundations of Research (pp. 59-63); Cambridge, Massachusetts: The MIT Press; quotations in text from reprinted version).

Latour, B. (1987) Science in Action: How to Follow Scientists and Engineers throughout Society . Milton Keynes, England: Open University Press.

Laudan, L. (1977) Progress and its Problems: Towards a Theory of Scientific Growth . Berkeley, California: University of California Press.

Law, J., & Barnes, B. (1976) Research note: areas of ignorance in normal science: a note on Mulkay's 'Three models of scientific development'. The Sociological Review, 24 , 115-124.

le Cun, Y. (1985) Un procedure d'apprentissage pour reseau a seuil assymetrique (A learning procedure for assymetric threshold network). Proceedings of Cognitiva, 85 , 599–604.

le Cun, Y. (1988) A theoretical framework for back–propagation. In D. S. Touretzky, G. E. Hinton, & T. J. Sejnowski (Ed.), Proceedings of the 1988 Connectionist Models Summer School San Mateo, California: Morgan Kaufmann, Inc.

le Cun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989) Backpropagation applied to handwritten zip code recognition. Neural Computation, 1 , 541-551.

Lemaine, G., MacLeod, M., Mulkay, M., & Weingart, P. (1976) Perspectives on the Emergence of Scientific Disciplines . The Hague: Mouton & Co.

Lettvin, J. (1988) Foreword. In W. S. McCulloch (Ed.), Embodiments of Mind (pp. v-xi). Cambridge, Massachusetts: The MIT Press.

Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. (1959) What the frog's eye tells the frog's brain. Proceedings of the IRE, 47 , 1940-1951.

Levin, J. A. (1976). Proteus: an activation framework for cognitive process models (ISI/WP–2). University of Southern California, Information Sciences Institute, Marina del Rey, California.

Lighthill, J. (1973) Artificial Intelligence . London: Science Research Council (Great Britain).

Lippmann, R. P. (1987) An introduction to computing with neural nets. IEEE ASSP Magazine, 4 April, 4-22.

Lorentz, G. G. (1976) The 13th problem of Hilbert. In F. E. Browder (Ed.), Mathematical Developments Arising from Hilbert Problems Providence, Rhode Island: American Mathematical Society.

Lowever, B., & Rey, G. (1991, in press) Fodor and His Critics . Oxford, England: Blackwell.

Lucky, R. W. (1965) Automatic equalization for digital communication. Bell Syst. Tech. J., 44 , 547-588.

Lucky, R. W., & et. al. (1968) Principles of Data Communication . New York: McGraw-Hill.

MacKenzie, D. (1990) Inventing Accuracy: A Historical Sociology of Nuclear Missile Guidance . Cambridge, Massachusetts: The MIT Press.

MacKenzie, D., & Wajcman, J. (1985) The Social Shaping of Technology . Milton Keynes, England: Open University Press.

Maddox, J. (1987) Modelling for its own sake. Nature, 328 , 571.

Marr, D. (1969) A theory of cerebellar cortex. Journal of Physiology (London), 202 , 437-470.

Marr, D. (1970) A theory for cerebral neocortex. Proceedings of the Royal Society of London, B 176 , 161-234.

Marr, D. (1971) Simple memory: A theory for archicortex. Philosophical Transactions of the Royal Society of London, B 262 841, 23.

Marr, D. (1982) Vision: A Computational Investigation into the Human Representation and Processing of Visual Information . New York: W. H. Freeman and Company.

McClelland, J. L. (1989) Interview. (Carnegie-Mellon University, Pittsburgh, Pennsylvania, 1/11/89, phone interview).

McClelland, J. L., & Rumelhart, D. E. (1981) An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. Psychological Review, 88 , 375-407.

McClelland, J. L., & Rumelhart, D. E. (1985) Distributed memory and the representation of general and specific information. Journal of Experimental Psychology: General, 114 2, 159-188.

McClelland, J. L., & Rumelhart, D. E. (1989) Explorations in Parallel Distributed Processing: A Handbook of Models, Programs, and Exercises . Cambridge, Massachusetts: The MIT Press.

McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. (1986) The appeal of parallel distributed processing. In D. E. Rumelhart, J. L. McClelland, & The-PDP-Research-Group (Ed.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 1, Foundations (pp. 3-44). Cambridge, Massachusetts: The MIT Press.

McClelland, J. L., Rumelhart, D. E., & The–PDP–Research–Group. (1986) Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 2, Psychological and Biological Models . Cambridge, Massachusetts: The MIT Press.

McCorduck, P. (1979) Machines Who Think: A personal Inquiry into the History and Prospects of Artificial Intelligence . New York: W. H. Freeman and Company.

McCulloch, W. S. (1965/1988) Embodiments of Mind . Cambridge, Massachusetts: The MIT Press (originally published in 1965; new edition in 1988).

McCulloch, W. S., & Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 5 , 115-133 (Reprinted in :Anderson, J. A., & Rosenfeld, E. (1988) (Ed.), Neurocomputing: Foundations of Research, (pp. 18-27); Cambridge, Massachusetts: The MIT Press; quotations in text from reprinted version).

McKenna, T. (interview) Interview. (Office of Naval Research, Washington DC, 21 November 1989).

Mead, C. (1989) Analog VLSI Systems and Neural Networks . Reading, Massachusetts: Addison-Wesley Publishing Company.

Mead, C., & Conway, L. (1980) Introduction to VLSI Systems . Reading, Massachusetts: Addison-Wesley Publishing Company.

Merton, R. K. (1942) Science and technology in a democratic order. Journal of Legal and Political Sociology, 1 , 115-126 (reprinted in R. K. Merton 1973 The Sociology of Science: Theoretical and Empirical Investigations. Chicago, Illinois: The University of Chicago Press, pp. 267-278. Quotations in text from reprinted version).

Merton, R. K. (1961) Singletons and multiples in science. Proceedings of the American Philosophical Society, 105 5, 470-486 (reprinted in Merton, R. K.1973 The Sociology of Science: Theoretical and Empirical Investigations; Chicago, Illinois: The University of Chicago Press, pp.343-370; quotations in text from reprinted version).

Merton, R. K. (1973) The Sociology of Science: Theoretical and Empirical Investigations . Chicago, Illinois: The University of Chicago Press.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960) Plans and the Structure of Behavior . New York: Holt, Rinehart, & Winston.

Minsky, M. L. (1954) Neural nets and the brain-model problem . Unpublished PhD dissertation, Princeton University.

Minsky, M. L. (1956) Some universal elements for finite automata. In C. E. Shannon, & J. McCarthy (Ed.), Automata Studies (pp. 117-128). Princeton: Princeton University Press.

Minsky, M. L. (1961) Steps towards artificial intelligence. Proceedings of the IRE, 49 January, 8-30.

Minsky, M. L. (1967) Computation: Finite and Infinite Machines . New York: Prentice-Hall.

Minsky, M. L. (1968) Introduction. In M. L. Minsky (Ed.), Semantic Information Processing (pp. 1-32). Cambridge, Massachusetts: The MIT Press.

Minsky, M. L. (1968b) Semantic Information Processing . Cambridge, Massachusetts: The MIT Press.

Minsky, M. L. (1987) The Society of Mind . London: Heineman.

Minsky, M. L. (interview) Interview. (MIT, Cambridge, Massachusetts, 25/11/89),

Minsky, M. L., & Papert, S. A. (1969) Perceptrons: An Introduction to Computational Geometry . Cambridge, Massachusetts: The MIT Press.

Minsky, M. L., & Papert, S. A. (1988) Perceptrons: An Introduction to Computational Geometry (expanded ed.). Cambridge, Massachusetts: The MIT Press.

Molina, A. H. (1987) The socio-technical basis of the microelectronics revolution: a global perspective. PhD, University of Edinburgh, Scotland.

Molina, A. H. (1990) Emerging neural computing in the USA, Japan, and UK/Europe. Science and Public Policy, 17 6, 363-371.

Mulkay, M. J. (1975) Three models of scientific development. Sociological Review, 23 , 509-526.

Mulkay, M. J., Gilbert, G. N., & Woolgar, S. (1975) Problem areas and research networks in science. Sociology, 9 , 187-203.

Nadel, L., Cooper, L. A., Culicover, P., & Harnish, R. M. (1989) Neural Connections, Mental Computation . Cambridge, Massachusetts: The MIT press.

New-York-Times. (1958a) New Navy device learns by doing. New York Times, July 8, 25:2.

New-York-Times. (1958b) Electronic 'brain' teaches itself. New York Times, July 13, iv9:6.
New-York-Times. (1971) Dr. Frank Rosenblatt dies at 43: taught neurobiology at Cornell. New York Times, July 13.
Newell, A. (1980) Physical symbol systems. Cognitive Science, 4, 135-183.
Newell, A. (1981). The knowledge level (CMU–CS–81–131). Carnegie–Mellon University.
Newell, A. (1983) Intellectual issues in the history of artificial intelligence. In F. Machlup, & U. Mansfield (Ed.), The Study of Information: Interdisciplinary Messages (pp. 187-227). New York: John Wiley & Sons.
Newell, A., & Simon, H. A. (1976) Computer science as empirical enquiry: symbols and search. Communications of the Association for Computing Machinery, 19, 113-126 (reprinted in J. Haugeland (Ed.) 1981, Mind Design, pp.35-66; Cambridge, Massachusetts: The MIT Press; quotations in text from reprinted version).
Newsweek. (1958) Human brains replaced? Newsweek, July 21, 50.
Nilsson, N. J. (1965) The Mathematical Foundations of Learning Machines. San Mateo, California: Morgan Kauffmann Publishers.
Nilsson, N. J. (1990) The Mathematical Foundations of Learning Machines. San Mateo, California: Morgan Kaufmann (with introduction by T. J. Sejnowski and H. White).
Nilsson, N. J. (interview) Interview. (Stanford University, Computer Science Department, Stanford, California, 3/11/89, phone interview).
Nilsson, N. J., & Raphael, B. (1967) Preliminary design of an intelligent robot. In J. T. Tou (Ed.), Computer and Information Sciences-2 (pp. 235-259). New York: Academic Press.
Norberg, A. L. (1990approx.). Government support for the development of new technology: the case of DARPA and computer science and engineering, 1962-1982. Charles Babbage Institute, University of Minnesota (Typescript, n.d.).
Norman, D. A. (interview) Interview. (UCSD, La Jolla, California, 8/11/89).
Norman, D. A., & Bobrow, D. G. (1975) On data-limited and resource-limited processes. Cognitive Psychology, 7, 44-64.
Norman, D. A., & Bobrow, D. G. (1976) On the role of active memory processes in perception and cognition. In C. N. Cofer

(Ed.), The Structure of Human Memory San Francisco: Freeman.

Norman, D. A., & Bobrow, D. G. (1979) Descriptions: an intermediate stage in memory retrieval. Cognitive Psychology, 11 , 107-123.

Norman, N. A. (1986) Reflections on cognition and parallel distributed processing. In J. L. McClelland, D. E. Rumelhart, & The-PDP-Research-Group (Ed.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 2, Psychological and Biological Models (pp. 531-546). Cambridge, Massachusetts: The MIT Press.

NPL. (1959) Mechanisation of Thought Processes (volumes I and II) . London: Her Majesty's Stationery Office.

Oaksford, M., Chater, N., & Stenning, K. (1990) Connectionism, classical cognitive science and experimental psychology. AI and Society, 4 , 73-90.

Papert, S. (1965) Introduction. In W. S. McCulloch (Ed.), Embodiments of Mind Cambridge, Massachusetts: The MIT press (new edition 1988).

Papert, S. A. (1988) One AI or many? In S. R. Graubard (Ed.), The Artificial Intelligence Debate: False Starts, Real Foundations (pp. 1-14). Cambridge, Massachusetts: The MIT Press.

Papert, S. A. (interview) Interview . (Massachusetts Institute of Technology, Cambridge, Massachusetts, 4/12/89).

Parker, D. B. (1985). Learning-logic (TR-47). Center for Computational Research in Economics and Management Science, MIT, Cambridge, Massachusetts.

Patridge, D., & Wilks, Y. (1990) The Foundations of Artificial Intelligence . Cambridge, England: Cambridge University Press.

Peláez, E. (1988). Parallelism and the crisis of von Neumann computing (Edinburgh PICT Working paper no. 5). Programme on Information and Communication Technology, University of Edinburgh, Research Centre for Social Sciences.

Pickering, A. (1981) Constraints on controversy: the case of the magnetic monopole. Social Studies of Science, 11 , 63-93.

Pinch, T. J. (1981) The sun-set: the presentation of certainty in scientific life. Social Studies of Science, 11 , 131-158.

Pinch, T. J., & Bijker, W. E. (1987) The social construction of facts and artifacts: on how the sociology of science and the sociology of technology might benefit each other. In W. E. Bijker, T. P. Hughes, & T. Pinch (Ed.), The Social Construction of Technological Systems: New Directions in the Sociology

and History of Technology (pp. 17-50). Cambridge, Massachusetts: The MIT Press.

Pinker, S., & Prince, A. (1988) On language and connectionism: analysis of a parallel distributed model of language acquisition. Cognition, 28 , 73-193.

Posner, M. I. (1978) Chronometric Explorations of Mind . Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Pratt, V. (1987) Thinking Machines: The Evolution of Artificial Intelligence . Oxford, England: Basil Blackwell.

Pylyshyn, Z. W. (1980) Computation and cognition: issues in the foundations of cognitive science. Behavioral and Brain Sciences, 3 , 111-169.

Pylyshyn, Z. W. (1984) Computation and Cognition: Toward a Foundation for Cognitive Science . Cambridge, Massachusetts: The MIT Press.

Qian, N., & Sejnowski, T. C. (1988) Predicting the secondary structure of globular proteins using neural network models. Journal of Molecular Biology, 202 , 865-884.

Quillian, M. R. (1968) Semantic memory. In M. L. Minsky (Ed.), Semantic Information Processing (pp. 216-270). Cambridge, Massachusetts: The MIT Press.

Raphael, B. (1976) The Thinking Computer: Mind Inside Matter . San Francisco: Freeman.

Rappa, M. A., & Debackere, K. (1989). The emergence of a new technology: the case of neural networks (WP#3031-89-BPS). Massachusetts Institute of Technology, Alfred P. Sloan School of Management.

Rappa, M. A., & Debackere, K. (1990). International survey on the neural network research community . Massachusetts Institute of Technology, Alfred P. Sloan School of Management.

Reddy, D. R., Erman, L. D., Fennell, R. D., & Neely, R. B. (1973) Hearsay speech understanding system: an example of the recognition process. Proceedings of the International Conference on Artificial Intelligence (pp. 185–194).

Ridgway–III, W. C. (1962) An adaptive logic system with generalizing properties . PhD thesis, tech. rep.1556-1, Stanford Electron. Labs., Stanford, California.

Robbins, H., & Monro, S. (1951) A stochastic approximation method. Annals of Math. Stat., 22 , 400-407.

Roberts, L. G. (1963). Machine perception of three dimensional solids (Technical report no. 315). MIT Lincoln Laboratory, Lexington, Massachusetts.

Rochester, N., Holland, J. H., Haibt, L. H., & Duda, W. L. (1956) Tests on a cell assembly theory of the action of the brain, using a large digital computer. IRE Transactions on Information Theory, IT-2 , 80-93.

Rosen, C. A. (interview) Interview . (Atherton, California, 10/11/89).

Rosenblatt, F. (1957). The perceptron, a perceiving and recognizing automaton (Project PARA) (85-460-1). Cornell Aeronautical Laboratory.

Rosenblatt, F. (1958a) The perceptron: a probabilistic model for information storage and organization in the brain. Psychological Review, 65 , 386-408 (Reprinted in J. A. Anderson & E. Rosenfeld (Ed.) (1988), pp. 92-114; quotations in text from reprinted version).

Rosenblatt, F. (1958b). The perceptron: a theory of statistical separability in cognitive systems (VG-1196-G-1). Cornell Aeronautical Laboratory, Buffalo, New York.

Rosenblatt, F. (1959) Two theorems of statistical separability in the perceptron. Mechanisation of Thought Processes (pp. 421-456). London: Her Majesty Stationery Office (Proceedings of a Symposium held at the National Physical Laboratory, November 1958, vol. 1).

Rosenblatt, F. (1960). On the convergence of reinforcement procedures in simple perceptrons (VG-1196-G-4). Cornell Aeronautical Laboratory, Buffalo, New York.

Rosenblatt, F. (1962a) Principles of Neurodynamics . New York: Spartan.

Rosenblatt, F. (1962b) Strategic approaches to the study of brain models. In H. von Foerster, & G. W. Zopf (Ed.), Illinois Symposium on Principles of Self-organization (University of Illinois, Urbana, Illinois) (pp. 385-396). New York: Pergamon Press.

Rosenblatt, F. (1964) A model for experimental storage in neural networks. In J. T. Tou, & R. H. Wilcox (Ed.), Computer and Information Sciences (pp. 16-66). Washington, DC: Spartan Books.

Rosenblatt, F. (1967) Recent work on theoretical models of biological memory. In J. T. Tou (Ed.), Computer and Information Sciences-II New York: Academic Press. (Proceedings of the Second Symposium on Computer and Information Sciences held at Batelle Memorial Institute).

Rosenblatt, F., Farrow, J. T., & Herblin, W. F. (1966) Transfer of conditioned responses from trained rats to untrained rats by means of a brain extract. Nature, 209 , 46-48.

Rosenblueth, A., Wiener, N., & Bigelow, J. (1943) Behavior, teleology, and purpose. Philosophy of Science, 10 , 18-24.
Rumelhart, D. E. (interview) Interview . (Stanford University, Stanford, California, 13/11/89).
Rumelhart, D. E., Hinton, G. O., & Williams, R. J. (1986) Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & The–PDP–Research–Group (Ed.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 1. Foundations (pp. 318-362). Cambridge, Massachusetts: The MIT Press.
Rumelhart, D. E., & McClelland, J. E. (1982) An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. Psychological Review, 89 , 60-94.
Rumelhart, D. E., & McClelland, J. L. (1985) Levels Indeed! A Response to Broadbent. Journal of Experimental Psychology: General, 114 2, 193–197.
Rumelhart, D. E., & McClelland, J. L. (1986a) PDP models and general issues in cognitive science. In D. E. Rumelhart, J. L. McClelland, & the-PDP-Research-Group (Ed.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 1. Foundations (pp. 110–146). Cambridge, Massachusetts: The MIT Press.
Rumelhart, D. E., & McClelland, J. L. (1986b) On learning the past tenses of English verbs. In J. L. McClelland, D. E. Rumelhart, & The-PDP-Research-Group (Ed.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 2. Psychological and Biological Models (pp. 216-271). Cambridge, Massachusetts: The MIT Press.
Rumelhart, D. E., & McClelland, J. L. (1986c) Future directions. In J. L. McClelland, D. E. Rumelhart, & The-PDP-Research-Group (Ed.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 2. Psychological and Biological Models (pp. 547-552). Cambridge, Massachusetts: The MIT Press.
Rumelhart, D. E., McClelland, J. L., & The–PDP–Research–Group. (1986) Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 1. Foundations . Cambridge, Massachusetts: The MIT Press.
Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. E. (1986) Schemata and sequential thought processes in PDP models. In J. L. McClelland, D. E. Rumelhart, & The-PDP-Research-Group (Ed.), Parallel Distributed Processing:

Explorations in the Microstructure of Cognition, vol. 2, Psychological and Biological Models (pp. 7-57). Cambridge, Massachusetts: The MIT Press.

Rumelhart, D. E., & Zipser, D. (1985) Feature discovery by competitive learning. Cognitive Science, 9, 75-112.

Rumelhart, D. J., Hinton, G. E., & Williams, R. J. (1986b) Learning representations by back-propagating errors. Nature, 323, 533-536 (reprinted in Anderson & Rosenfeld (Ed.) 1988, pp. 696-699, quotations in text from reprinted version).

Schon, D. A. (1963) Invention and the Evolution of Ideas. London: Social Science Paperbacks with Tavistock Publications.

Schwartz, T. J. (1988) 1987, the neural year in review. Synapse Connection (now Neural Technology Update), 2 2 (February), 1 and 11-12.

Schwartz, T. J. (1989) 1988, the neural network year in review. Synapse Connection (now Neural Technology Update), 3 1 (January), 1 and 12-13.

Sejnowski, T. J. (1987) Computing with connections. Journal of Mathematical Psychology, 31, 203-210.

Sejnowski, T. J., Koch, C., & Churchland, P. S. (1988) Computational neuroscience. Science, 241, 1299-1306.

Sejnowski, T. J., & Rosenberg, C. R. (1986). NETtalk: a parallel network that learns to read aloud (JHU/EECS-86/01). The John Hopkins University Electrical Engineering and Computer Science (reprinted in Anderson & Rosenfled (Ed.) 1988, pp.663-672, quotations in text from reprinted version).

Sejnowski, T. J., & Rosenberg, C. R. (1987) Parallel networks that learn to pronounce English text. Complex Systems, 1, 145-168.

Shapin, S. (1979) The politics of observation: cerebral anatomy and social interests in the Edinburgh phrenology disputes. The Sociological Review Monograph, 27 March, 139-178.

Shapin, S. (1982) History of science and its sociological reconstructions. History of Science, 20, 157-211.

Sholl, D. A. (1956) Organization of the Cerebral Cortex. London: Methuen & Co.

Shondi, M. M. (1967) An adaptive echo canceller. Bell Syst. Tech. J., 46, 497-511.

Smolensky, P. (1987) The constituent structure of connectionist mental states: a reply to Fodor and Pylyshyn. The Southern Journal of Philosophy, 26 supplement, 137-163.

Smolensky, P. (1988) On the proper treatment of connectionism. The Behavioral and Brain Sciences, 11, 1-74.

Smolensky, P. (interview) Interview . (Denver Airport, Colorado, 14/11/89).

Star, S. L. (1989a) Regions of the Mind: Brain Research and the Quest for Scientific Certainty . Stanford, California: Stanford University Press.

Star, S. L. (1989b) The structure of ill-structured solutions: boundary objects and heterogeneous distributed problem solving . Manuscript (University of California, Department of Computer Science, Irvine, California), to appear in: M. Huhns & Les Gasser (Eds.), Readings in Distributed Artificial Intelligence 2, Morgan Kauffman, Menlo Park, California.

Steinbuch, K. (1961) Die Lernmatrix. Kybernetik, 1 , 36-45.

Sutherland, S. (1986) Parallel distributed processing. Nature, 323 9, 486.

Swaine, M. (1989) Parker's perceptions. Dr. Dobb's Journal, October, 112-121.

Swaine, M. (1989b) Two early neural net implementations. Dr. Dobb's Journal, 14 11, 124-131.

Tangney, J. (interview) Interview . (Air Force Office for Scientific Research, Washington DC, 21/11/89).

Taylor, W. K. (1956) Electrical simulation of some nervous system functional activities. In E. C. Cherry (Ed.), Information Theory (pp. 3). London: Butterworths.

Taylor, W. K. (1959) Pattern recognition by means of analogous automatic apparatus. Proceedings of the IEE London, 106B , 168-172.

Tesauro, G. (1990) Neurogammon wins computer olympiad. Neural Computation, 1 , 321-323.

The-Econominst. (1987) What the brain builders have in mind. The Economist, May 2, 94-96.

The-New-Yorker. (1958) Rival. The New Yorker, December 6, 44-45.

von der Malsburg, C. (1973) Self-organization of orientation sensitive cells in the striata cortex. Kybernetik, 14 , 85-100.

von der Malsburg, C. (1986) Frank Rosenblatt: Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. In G. Palm, & A. Aersten (Ed.), Brain Theory (pp. 245-248). Berlin: Springer-Verlag.

von der Malsburg, C. (interview) Interview . (University of Southern California, Los Angeles, 7/11/89).

von der Malsburg, C., & Willshaw, D. J. (1977) How to label nerve cells so that they can interconnect. Proceedings of the National Academy of Sciences USA, 74 11, 5176-5178.

von Foerster, H., & Zopf, G. W. (1962) Illinois Symposium on Principles of Self-organization (University of Illinois, Urbana, Illinois) . New York: Pergamon Press.

von Neumann, J. (1951) The general and logical theory of automata. In L. A. Jeffress (Ed.), Cerebral Mechanisms in Behavior. The Hixon Symposium (pp. 1-31). New York: John Wiley & Sons Inc.

von Neumann, J. (1956) Probabilistic logics and the synthesis of reliable organisms from unreliable components. In C. E. Shannon, & J. McCarthy (Ed.), Automata Studies (pp. 43-98). Princeton: Princeton University Press.

von Neumann, J. (1958) The Computer and the Brain . New Haven: Yale University Press.

Werbos, P. J. (1974) Beyond regression: new tools for prediction and analysis in the behavioral sciences. PhD, Harvard University, Cambridge, Massachusetts.

Werbos, P. J. (1982) Applications of advances in nonlinear sensitivity analysis. In R. F. Drenick, & F. Kozin (Ed.), Systems Modelling and Optimization: Proceedings of the 10th IFIP Conference, New York City, USA, August 31-September 4, 1981 New York: Springer-Verlag.

Werbos, P. J. (1988) Generalization of backpropagation with application to a recurrent gas market model. Neural Networks, 1 , 339-356.

Werbos, P. J. (interview) Interview . (National Science Foundation, Washington DC, 2/11/89).

White, H. (1989) Learning in artificial neural networks: a statistical perspective. Neural Computation, 1 , 425-464.

Widrow, B. (1960b). An adaptive 'adaline' neuron using chemical 'memistors' (Tech. rep. 1553-2). Stanford Electron. Labs.

Widrow, B. (1962) Generalization and information storage in networks of adaline "neurons". In M. C. Yovits, G. T. Jacobi, & G. D. Goldstein (Ed.), Self-Organizing Systems-1962 (pp. 435-461). Washington, DC: Spartan Books.

Widrow, B. (interview) Interview . (Electrical Engineering, Stanford University, Stanford, California, 13/11/89).

Widrow, B., & Hoff, M. E. (1960) Adaptive switching circuits. 1960 IRE WESCON Convention Record (pp. 96-104). New York: IRE.

Widrow, B., & Lehr, M. A. (1990) 30 years of adaptive neural networks: perceptron, madaline, and backpropagation. Proceedings of the IEEE, 78 9, 1415-1442.

Widrow, B., Mantey, P., Griffiths, L., & Goode, B. (1967) Adaptive antenna systems. Proceedings of the IEEE, 55 , 2143-2159.

Widrow, B., & Stearns, S. D. (1985) Adaptive Signal Processing . Englewood Cliffs, New Jersey: Prentice-Hall.

Wiener, N. (1948) Cybernetics. Scientific American, November , 14-18.

Wightman, C. W. (1959). Project PARA technical memorandum No. 4 . Cornell Aeronautical Laboratory, Buffalo, New York.

Wilks, Y. (1990) Some comments on Smolensky and Fodor. In D. Patridge, & Y. Wilks (Ed.), The Foundations of Artificial Intelligence: A Sourcebook (pp. 327–336). Cambridge, England: Cambridge University Press.

Will, C. A. (1989) Neural networks for defense: a conference report. Synapse Connection, 3 6, 12-13.

Williams, R. J. (interview) Interview . (Northeastern University, Boston, Massachusetts, 3/11/89).

Willshaw, D. J. (1971) Models of distributed associative memory . Unpublished PhD dissertation, University of Edinburgh, Scotland.

Willshaw, D. J. (1981) Holography, associative memory, and inductive generalization. In G. E. Hinton, & J. A. Anderson (Ed.), Parallel Models of Associative Memory (pp. 83-104). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Willshaw, D. J., Buneman, O. P., & Longuet-Higgins, H. C. (1969) Non-holographic associative memory. Nature, 222 , 960-962.

Willshaw, D. J., & von der Malsburg, C. (1979) A marker induction mechanism for the establishment of ordered neural mappings: its application to the retinotectal problem. Philosophical Transactions of the Royal Society of London, 287 , 203-243.

Yoon, B. (interview) Interview . Defense Advanced Research Projects Agency (DARPA), Arlington, Virginia, 20/11/89).

Yoon, B. L. (1989a). Artificial neural network technology . Defense Advanced Research Projects Agency (DARPA) 15 Feb.

Yovits, M. C. (interview) Interview . (Purdue University, Department of Computer and Information Science, Indianapolis, Indiana, 28/11/89, phone interview).

Yovits, M. C., & Cameron, S. (1960) Self-Organizing Systems: Proceedings of an Interdisciplinary Conference (Chicago 5–6 May 1959) . New York: Pergamon Press.

Yovits, M. C., Jacobi, G. T., & Goldstein, G. D. (1962) Self-organizing Systems 1962 . Washington, DC: Spartan.

Zeitvogel, R. K. (1988a) ICNN reviewed. Synapse Connection (now Neural Technology Update), 2 8, 10-11.

Zeitvogel, R. K. (1988b) INNS: the society for neural networking. Synapse Connection (now Neural Technology Update), 2 8, 1 and 12.

◆ Errata

Page 12, line 3. Says: 'significantly in) certain classification tasks.' → should say: 'significantly in certain classification tasks).'.
Page 15, line 18. Says: 'Fifth generation' → should say: 'Fifth Generation'.
Page 20, line 7. Says: 'Pitt's' → should say: 'Pitts' '.
Page 20, line 10. Says: 'Pitt's' → should say: 'Pitts' '.
Page 21, line 24. Says: 'feedback and control' → should say: 'feedback and homeostasis'.
Page 58, line 10. Says: 'sum of activation' → should say: 'sum of the activation'.
Page 69, line 1. Says: 'amount of error made, whereas in the perceptron' → should say: 'in the perceptron'.
Page 77, line 9. Says: 'different way' → should say: 'different way of'.
Page 104, line 28. Says: 'p. 235) and,' → should say: 'p. 235), and'.
Page 109, line 27. Says: 'of perceptron' → should say: 'of the perceptron'.
Page 124, line 1. Says: 'functions, and exclusive-or is not linearly separable. This issue' → should say: 'This issue'.
Page 176, line 27. Says: 'undoubtedly though' → should say: 'undoubtedly thought'.
Page 178, line 10. Says: 'in the important paper' → should say: 'in their important paper'.
Page 182, line 18. Says: 'in quotation' → should say: 'in the quotation'.
Page 191, line 31. Says: 'emphasis if laid' → should say: 'emphasis is laid'.
Page 195, line 1. Says: 'early 1980s' → should say: 'early 1970s'.
Page 199, line 12. Says: 'or model using' → should say: 'or modelled using'.

Page 201, line 7. Says: 'perception cognition' → should say: 'perception and cognition'.
Page 202, line 13. Says: 'approach in apparent' → should say: 'approach is apparent'.
Page 202, line 15. Says: 'on to the association' → should say: 'on the association'.
Page 207, line 22. Says: '— were' → should say: '— was'.
Page 230, line 8. Says: 'Hopfield-like with' → should say: 'Hopfield-like network with'.
Page 246, line 2. Says: 'at all to' → should say: 'at all to the'.
Page 276, line 4. Says: ' 'crystalising' ' → should say: ' 'crystallising' '.
Page 289, line 24. Says: '(ibid., p. 368)' → should say: '(Molina, p. 368)'
Page 312, line 4. Says: ' 'rhetorical tactics' ' → should say: 'rhetorical tactics.' '.
Page 315, line 20. Says: 'had problems' → should say: 'had had problems'.
Page 317, line 30. Says: '(5.4:' → should say: '(section 5.4:'.
Page 322, line 14. Says: 'statistical systems' → should say: 'statistical physics systems'.
Page 322, line 20. Says: 'The are about' → should say: 'They are about'.