A Sociological History of the Neural Network Controversy

MIKEL OLAZARAN

Universidad Pública de Navarra
Departamento de Sociología
Campus de Arrosadía
Pamplona, Spain

1. Introduction: A Sociological View of Scientific Controversies

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) that 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

ISBN 0-12-012137-9

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 differs from the tradition that 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 these types of tasks.

In this chapter, I study the scientific controversies that have shaped neural network research from a sociological point of view. The sociology of science and technology deals with the social processes—both internal and external to the research community—through which scientific knowledge and technological systems are generated and validated. Of course by saying that science is produced and assessed socially I do not mean that it is inadequate or "ideologically bad." The sociology of science claims that all science, whether it is seen as "good" or "bad", is socially generated and evaluated.

Science and technology are often generated and validated through debates and controversies (Collins, 1985; Latour, 1987; Star, 1989). In controversies the "interpretative flexibility" (Collins, 1981) of scientific results (data, experiments, findings) is more evident than in periods of consensus. By "interpretative flexibility" I mean, using Donald MacKenzie's terms, that "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 showing the interpretative flexibility of scientific knowledge is only the first step in a sociological study of science. The second one is to study the processes through which that interpretative flexibility, which *in principle* could always go on, is *in practice* brought to closure. The sociology of science claims that social factors—both internal to the research community and involving the wider society—are at the basis of those processes of closure.

> Why some knowledge claims are challenged and why some are not, and why some challenges succeed and some fail, thus become interesting empirical questions. Central to the answers are matters of the interests, goals, traditions, and experiences of the social groups ([scientific], technological and other) involved; of the conventions surrounding technological testing; and of the relative prestige and credibility of different links in the network of knowledge. (MacKenzie, 1990, pp. 10–11)

This "controversy-closure" scheme is especially interesting in the case of the history of neural networks, which has been shaped by controversies. They were quite popular in the late 1950s and early 1960s, but were almost abandoned in the second half of the 1960s when the "perceptron controversy" was closed and symbolic AI emerged as "the right approach" to AI. However, almost 20 years later, in a different context, the neural network controversy reopened, and many of the conclusions of the early debate were revised and changed. Although this second controversy has not been closed yet, it has already brought about the acceptance of neural networks as an approach to AI in its own right.

Debate is a positive force in the generation of science and technology, which develop through processes of controversy and closure. In scientific controversies, participants make use of many types of rhetorical (or debating) tactics. Rhetoric is always used in the processes of closure of scientific controversies, and it is therefore an element of scientific activity. But I would like to emphasize that when I look at the rhetoric used in the neural network controversies I am not criticizing any of the positions involved. My aim is to study the main developments of the history of neural networks, and I claim that those developments are best studied by looking at the neural network controversies. I do not aim at evaluating the positions involved; my objective is to show how controversies have shaped neural network research.

Here I analyze the main developments of the neural network controversy from a sociological point of view. In Section 2 I look at the controversy that surrounded Frank Rosenblatt's perceptron machine in the late 1950s and early 1960s. It has often been argued that Rosenblatt made exaggerated claims, and that he somehow "provoked" the reaction of symbol-processing researchers. However, in Section 3 I show that Rosenblatt was well aware of the main problems of his machine, and that he even insisted on them in his books and papers. Section 4 concentrates on one of the main problems of early neural network research, namely the issue of training multilayer systems.

In the middle of the perceptron controversy, Minsky and Papert embarked on a project aimed at showing the limitations of Rosenblatt's perceptron beyond doubt. In Section 5, I analyze some of the main results of that project, and I show that Minsky and Papert (on the one hand) and neural network researchers (on the other) interpreted those results rather differently (this is therefore a case of interpretive flexibility). In Sections 6 through 8, I discuss the processes through which this interpretative flexibility was closed and the effects that the crisis of early neural network research had upon the three most important neural network groups of the time, namely Widrow's group, Rosenblatt's group, and the group at SRI. In Section 8, I also look at the influence that factors like the emergence of symbolic AI and computer technology had on the closure of the neural network controversy.

After the closure of the perceptron controversy, symbol-processing remained the dominant approach to AI over the years, until the early 1980s. In Section 9, I review briefly some of the most important aspects of that changing context. In Section 10 I look at the history of back-propagation, and in particular at Werbos' unsuccessful attempts to sell this idea in the 1970s and early 1980s. Section 11 elaborates on back-propagation, one of the most successful neural network techniques of the second half of the 1980s. The success of back-propagation has to be understood within the context of re-emergence of neural network research in the 1980s. But the idea of back-propagation has an interesting history.

Neural networks research exploded in the late 1980s. In Section 12 some indicators of the growth of the neural network community are examined. The recent (re)emergence of neural networks in the late 1980s has brought about the reopening of the neural network controversy. In Section 13 I review some of the main positions that can be found in this new controversy, and I emphasize the open character of the current debate. Finally, at the end of Section 13 I reflect briefly on some of the main issues discussed throughout this chapter.

2. The Controversy of the Perceptron

During the 1950s, neural networks and symbolic AI emerged as the two main approaches to both studying cognition computationally (what today is called cognitive science) and building intelligent machines (today's artificial intelligence). Probably because of the early, undeveloped stage of both neural networks and symbol-processing AI at that time, these two approaches were seen by many as *alternative* (rather than *complementary*) solutions to the problems of those disciplines. In this and the following sections I analyze some of the main developments of the early phase of the history of neural network research. It is my view that those developments can be studied by looking closely at Frank Rosenblatt's perceptron machine.

The controversy of the perceptron exploded as the results of Rosenblatt's perceptron project received considerable attention outside the research community. Rosenblatt published his first important papers on the perceptron in 1958 (Rosenblatt, 1958a, 1958b). At that time, a team of researchers started to build the Mark 1 perceptron at Cornell Aeronautical Laboratory (CAL, Buffalo, New York, today Arvin Calspan Advanced Technology Center) funded by the Office of Naval Research (ONR). The perceptron was a feedforward neural network machine with one layer of adjustable connections. Rosenblatt developed a "learning" algorithm that guaranteed that the perceptron could learn any classification task that could be embodied by its structure of units and connections.

Figure 1 is a simplified representation of Rosenblatt's perceptron machine. The perceptron can be defined as a single-layer feedforward neural network. Here *single-layer* means that there is only one layer of modifiable connections (namely the connections from association units to response unit in Fig. 1). As it is shown in the figure, 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.

The perceptron built by Rosenblatt's group at CAL had eight response units, but only one of them is represented in Fig. 1. The response units had modifiable incoming connections. Units of this type have been the building block of many neural computing systems since. The response units of the perceptron were basically McCulloch and Pitts formal neurons with modifiable connections. A unit of this type fires if the sum of activation it receives

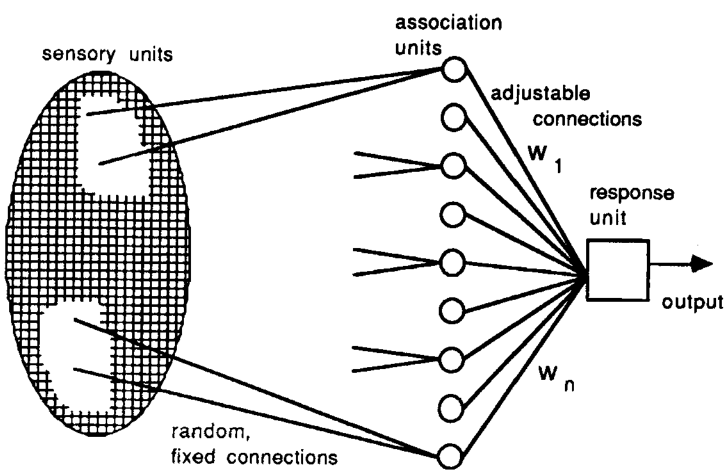

FIG. 1. Perceptron

from other units equals or exceeds its threshold value. Note that the activation that a unit receives through each of its input lines is multiplied by the value of those lines or connections.

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 Widrow and Hoff (1960). 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.[1]

The perceptron convergence theorem was proved for the simplified perceptron of Fig. 2 (representing the adjustable part of the original perceptron after removing the sensory-to-association fixed connections). The perceptron learning algorithm states that, for learning to occur, it is necessary that the perceptron architecture be capable of embodying the desired input/output classification. But proving whether a classification can be carried out by the simplified perceptron of Fig. 2 (let alone the Mark 1 perceptron with its first layer of randomly wired connections) is an 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 generalization abilities required after training are very important.

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*. Another important characteristic of the perceptron learning algorithm is that error is minimized for each output unit independently of the others. Widrow and Hoff's LMS algorithm was different in this respect. LMS minimized error as summed over all the output units.

[1] "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 ..." (Rosenblatt, 1962a, p. 111).

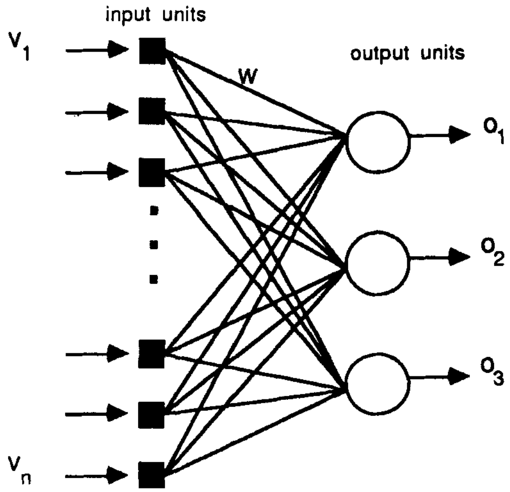

FIG. 2. Simplified perceptron.

Rosenblatt's perceptron received a lot of attention both inside and outside the scientific community. In July 1958, the Perceptron Project was announced at a press release held in Washington, D.C.. Marshall Yovits from ONR and Frank Rosenblatt himself participated in it. There, ONR announced its financial support for Rosenblatt's project, and Rosenblatt made a demonstration of the capabilities of the perceptron simulating it on a digital computer. As a consequence of that announcement, the perceptron project was widely reported in the press:

> 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. (*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)

Other examples of this type of press reporting can be found.[2]

The experiment reported by the *New York Times* in the quotation just given which says that the perceptron learns between right and left, is a good example of the capabilities and the problems of the perceptron. In this case, the perceptron was distinguishing between right and left, but it was not recognizing the squares as the same object (I will come to this point later). However, the fact that the perceptron learned something *at all* was presented as a success.

Much has been said about Rosenblatt's "exaggerated claims." In some widely held "popular" versions of the history of neural network research, these allegedly exaggerated claims have been given a prominent role. On occasions when a few comments on the history of neural networks are required, such as in short "historical" introductions to papers and books, it is not uncommon to find statements like the following:

> 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. (Cruz, 1988, p. 2, emphasis added)

According to this "popular" version of the history of neural networks, symbolic AI researchers' strong reaction against the perceptron and neural

[2] See, for example: *New York Times* (1958a, 1958b), *Newsweek* (1958), and *The New Yorker* (1958). 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). See also these statements: "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).

network research in general (I examine this reaction later) was "justified" in order to stop neural network "overselling." However, it is my view that this emphasis on overselling not only misses some very important points of the history of neural networks, but it could also be a reconstruction made by the "winning side" of the perceptron controversy once that controversy was closed.

It is important to stress that the statements made by Rosenblatt in the 1958 press release have to be understood within the context of legitimation of his research project *outside* the scientific community.[3] Presenting and publicizing the perceptron project was important for ONR too. They also, as a government organization, had to justify their research-supporting activity. Marshall Yovits, who was responsible for the funding of the perceptron project at ONR (Information Systems Branch) at the time, was also at the 1958 press release. When I interviewed him, he complained strongly about the reaction from symbolic AI researchers both to Rosenblatt's work and 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 organization—in order to continue to get

[3] Science is a social activity, and scientific knowledge the product of that activity. Every social activity has to be constantly sustained (i.e., it has a need of legitimation). The social legitimation of science can be studied at two (interrelated) levels, namely at the level of the scientific community and at the level of the wider society. Right now I am talking about this second level (however, in the rest of the chapter I make many considerations of the first one). It is obvious that, because of the economic costs of scientific research, there is always competition for resources and for legitimation in the wider society. Latour goes further to affirm that the bigger the "inside" of a research project, the bigger its "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. "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" (Latour, 1987, p. 156). The topic of the relationship between AI research and the wider society is especially interesting because there are many discourses about human intelligence and human action in society that 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).

> 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 publicize it, leading then to more support. (Yovits, interview)

Rosenblatt himself complained about the press reporting of the perceptron project.[4]

But let me come to symbol-processing researchers' reactions to Rosenblatt's claims.

> 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 [McCorduck 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' . . . (McCorduck, 1979, p. 87)

The rhetoric used by researchers both in favor of and against neural networks are a good indicator of the level of controversy reached. A frequent rhetorical move by Rosenblatt's critics was to accuse him of "irritating people."

> 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 heads does are inscrutable . . . Rosenblatt only irritated him. (McCorduck, 1979, p. 88)

But opinions were divided, and a significant number of researchers were convinced by Rosenblatt's arguments, charisma, and persuasion efforts.[5] A good example of this is Rosenblatt's visit to Stanford Research Institute

[4] ". . . 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 exuberance 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" (Rosenblatt, 1962a, p. v).

[5] ". . . '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' . . ." (McCorduck, 1979, p. 88)

(SRI). Sometime after this visit, Charles Rosen's group at SRI decided to start their Minos neural network project.

> 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)

The controversy of the perceptron was often personalized in the figures of Rosenblatt and Minsky. They were frequently the leaders of the two contending positions. Minsky and Rosenblatt engaged in heated debates at scientific conferences in the late 1950s and early 1960s.

> 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 adjustments 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] . . . 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)

> 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. (Rosenblatt, 1962a, p. v)

> 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 limited success, and to the conviction that this was not the route to advance the science of intelligent machines. (Rosen, interview)

I infer from my contacts with some of the researchers involved (see Appendix 1) that the atmosphere of the controversy was rather bitter at times, and that there were moments when diplomacy was left behind.

Minsky and Rosenblatt were often the leaders of the two main positions, but controversy extended to other neural network groups too. The SRI group is a good example of this.

> 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)

Competition for funding resources seems to have been one of the main reasons for this strong controversy. In the late 1950s, both neural networks and symbol-processing were emerging areas of research. None of them had reached consolidation by that time, and therefore competition for funding—and for scientific legitimacy in general—was particularly strong. Symbol-processing researchers saw neural network researchers' efforts to obtain funding for building more complex machines as a direct threat to their research interests. In this context, two of them—Marvin Minsky and Seymour Papert—decided to carry out a study that would show the limitations of neural networks beyond doubt. In later sections I will analyze Minsky and Papert's work and its importance in the closure of the perceptron controversy. But before that, I would like to turn my attention to some of the main problems that early neural network researchers were having with their machines.

3. The Problems of Early Neural Networks

In this section, I discuss briefly some of the problems that Rosenblatt was having with his perceptron machine in the late 1950s and early 1960s. I show that Rosenblatt and his colleagues were aware of the limitations of their single-layer neural network machine.

In the previous section, I said 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. There I anticipated the view that this "popular version" misses important aspects of the history of early neural network research. In Section 2, I also pointed out that the salience that the "overclaiming hypothesis" has gained over the years is probably a consequence of the closure of the perceptron controversy (to be analyzed later). Thus, it could have been the case that the accusation of overclaiming (a rhetorical element used by the neural network critics) became part of the "official" history of neural networks *after* (and not before) the controversy was closed against the neural network position.

Rosenblatt and his colleagues were aware of the problems of the perceptron, and acknowledged them openly in the scientific papers they published in the late 1950s and early 1960s. In this section, I will look at these problems in some detail.

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 openly admitted, that the perceptron did not classify objects according to their geometrical similarity (Rosenblatt, 1958a, p. 96, 1962a, pp. 67–70, 1962b, pp. 390–391). Instead, classifications were based on the amount of overlap or intersection between objects in the perceptron's input retina. If the amount of overlap between the retinal areas occupied by two objects was big enough, then the perceptron would be able to classify them under the same category (otherwise it would not).[6] Rosenblatt used the term "weak generalization" to refer to this type of overlap-based recognition, as opposed to "pure generalization," which the elementary perceptron was *not* capable of.

[6] For instance, the machine could, under the right circumstances, recognize 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 would not be classified as belonging to the same category ("square") by an elementary perceptron.

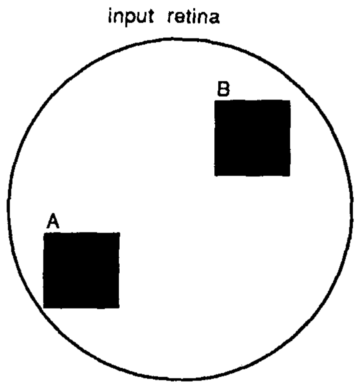

FIG. 3.

> ... 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 recognized them openly in his *Principles of Neurodynamics* book (e.g., Rosenblatt, 1962a, pp. 306–310). One of the problems, the issue of preprocessing (i.e., the problem of distinguishing the components of an image and the relationships between them), was related to the previously 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.[7] 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 (1962a, pp. 309–310) included the "figure-ground" or "connectedness" problem within this last point. It will be seen later that this question (i.e., the problem of recognizing a figure as distinct from its background) was one of the most important points of Minsky and Papert's (1969) arguments against single-layer perceptrons.

In dealing with the problem of recognizing similar objects appearing in different positions in the perceptron's retina, Rosenblatt studied systems with two layers of association units, and also systems with connections among the units of the same layer ("cross-coupled" perceptrons). He also carried out some research on "four-layer" perceptrons with one or more layers of modifiable connections (today's multilayer networks). Rosenblatt (1962a, p. 576) claimed that the perceptron's generalization capability improved considerably with these changes.[8] Nevertheless, he openly admitted that very important problems concerning four-layer (i.e., multilayer) and cross-coupled perceptron systems remained to be solved.

[7] "... 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).

[8] In a perceptron with two layers of association units ("four-layer" perceptrons in Rosenblatt's terms), the units of the first association layer, responded to similar features in different positions, would all activate the same unit in the second association layer, and in this way a feature in different positions could be recognized as the same (von der Malsburg, 1986, pp. 245–246).

Rosenblatt (1962a, pp. 577–579) summarized the limitations of perceptrons (both single-layer and more complex machines) in a list of 15 problems, some of which are reproduced 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. . . ." (Rosenblatt, 1962a, pp. 577–579)[9]

This quotation shows beyond doubt that Rosenblatt was well aware of the considerable difficulties faced by early neural network research. In points 4 and 7 in this quotation, the difficulties of training multilayer networks and, in particular, the lack of an effective algorithm for doing so, are clearly stated. Rosenblatt recognized that issues 4 and 7 were (in his words) "theoretical," meaning 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)

In point 13 of his list, Rosenblatt insists once again on the figure-ground or "connectedness" problem, one of the problems to which Minsky and Papert devoted a lot of attention in their critical study of perceptrons (Minsky and Papert, 1969). Rosenblatt's most pessimistic comments were for problems 13 (the figure-ground problem) and 14 (the recognition of topological relationships and abstract concepts).

[9] 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.

> 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 1st to the 12th] 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 and his colleague David Block insisted also on the problems found in recognizing topological and temporal relationships with the perceptron (Rosenblatt, 1958a, pp. 110–111, 1962b, pp. 390–391; Block, 1962, p. 149). These included predicates of the type "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 previous quotation) that progress toward solving these problems, as well as the figure-ground problem, had been almost insignificant.

Thus, Rosenblatt was well aware of the problems of the single-layer perceptron. He thought that further progress in neural network research would come from perceptrons more complex than the single-layer one, but he recognized that many problems regarding more complex perceptrons (including multilayer ones) remained to be solved.

4. Training Multilayer Networks: A "Reverse Salient" of Neural Network Research

In this section, I would like to look in more detail at one of the most important problems of early neural network research, namely training networks with more than one layer of modifiable connections (multilayer networks).

In this discussion of the problem of training multilayer networks and its importance in the crisis of early neural network research I will use Thomas Hughes' concepts of "reverse salient" and "critical problem" (Hughes, 1983). Reverse salients are problems that obstruct the development of technological systems.[10] According to Hughes, these problems are obvious to the agents involved in such a system. Therefore, the difficulty does not lie

[10] 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 (Hughes, 1983, 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.

in localizing a reverse salient, but in giving a satisfactory solution to it. When a reverse salient is defined as a problem that can be solved, it becomes a critical problem. For Hughes, defining reverse salients as critical problems is the key to technological innovation and change (Hughes, 1983, pp. 14–15, 22).

I showed earlier that the problem of learning in perceptrons with more than one layer of adjustable connections was seen by Rosenblatt and colleagues as one of the most important barriers to making further progress in perceptron research. With his convergence theorem, 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., noninfinite) number of repetitions of the presentation of input-adjustment of weights cycle. The problem was, of course, that there were some classification tasks that the single-layer perceptron could not realize. The typical example here is the *exclusive-or* function.

The problem of the exclusive-or logical function is 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 easy to see that this network can realize the AND function. Values for the parameters of the system (connection weights and the threshold of the output unit) can be found that embody this function. The AND function is 1 only when both inputs are 1, and it is 0 otherwise. So the device of Fig. 4 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.

The input space of the network of Fig. 4 can be represented as a two-dimensional space. The computation realized by the output unit (where weights and threshold intervene) separates that input space into two regions, one corresponding to output value 1 and the other one to output value 0. The AND function is linearly separable (a straight line that separates the two classes of outputs can be drawn).

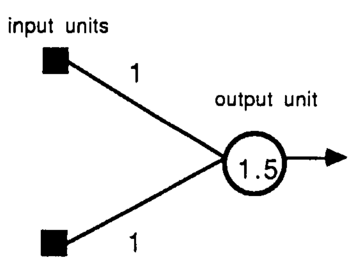

FIG. 4. AND function.

The exclusive-or function cannot be embodied by the system of Fig. 4. This system can only realize linearly separable functions, and exclusive-or is 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 that value. This problem can be solved by introducing an intermediate (or *hidden*) unit. 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.

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. The problem was that no weight modification rule had been developed for multilayer networks that guaranteed results comparable to those obtained with both the perceptron learning algorithm and Widrow and Hoff's LMS (least mean square) algorithm for single-layer networks. Neural network researchers were aware of the problem of training multilayer perceptrons *long before* Minsky and Papert's (1969) study was published. A good example of the early salience of this problem can be found in J. K. Hawkins' (1961) review about "self-organizing systems." In this review the issue of training multilayer networks figured prominently among the problems of neural networks.

> 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)

The problem of learning in multilayer systems figured prominently in the research agenda of the main neural network groups of the late 1950s and early 1960s. In the previous section, I showed that this problem was very important for Rosenblatt and his colleagues. In the rest of this section I will discuss how this issue affected both Widrow's group at Stanford University and Rosen's group at Stanford Research Institute (SRI).

Widrow and his colleagues were aware that their most powerful single-layer neural network machines had important problems and limitations. They were aware that there were important classification tasks that the single-layer Madaline could not realize, and that machines with more layers of modifiable connections had much greater classification power. They studied multilayer madalines and investigated some learning procedures for them. These were madalines in which the second layer of connections was also adaptable (they were what today is called multilayer networks). Figure 5 shows one type of multilayer architecture studied by Widrow and his colleagues.

The system represented in Fig. 5 is capable of classifying inputs into eight categories (it has three binary output units). In some of his experiments, Widrow used a 4 × 4 square retina (i.e., 16 input units that could represent letters and other patterns). He also studied systems with bigger input "retinas." The objective of Widrow's experiments with the multilayer network of Fig. 5 was:

> . . . To teach the system to classify . . . patterns [belonging to the eight 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

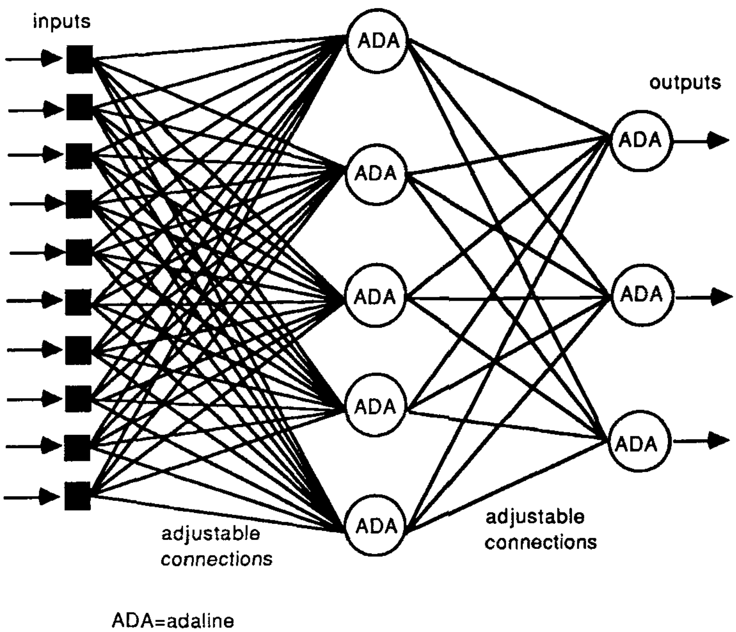

FIG. 5. Madaline with two layers of modifiable connections.

> 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 madalines 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. 455)

Widrow described a procedure for training a system like the one in Fig. 5 which "had been found by experiment to work well" (Widrow, 1962, p. 456). However, results could not be guaranteed in the same sense as Widrow and Hoff's (1960) LMS algorithm guaranteed results for single-layer machines. Widrow reported some "successful" experiments on "wide varieties of specific responses" with a small scale version of the system of Fig. 5, namely a system with three adalines in the first layer of units and one in the second (and therefore two layers of modifiable connections). He and his colleagues dedicated important research efforts to the problem of training multilayer machines.

> The above procedure [for adapting weights] 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)

But even though many experiments were carried out, Widrow and his colleagues were not able to develop a learning algorithm for multilayer networks comparable to the one they had earlier developed for single-layer networks. In part as a result of that, they started to shift their focus of attention to engineering applications of adalines and the LMS algorithm outside neural network research.

> We tried to adapt layered [i.e., multilayer] neural nets, but we never succeeded. 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, but we never succeeded. I couldn't imagine any way to do it. I think that—if anything—for lack of success on that problem I stopped working on neural networks and switched to other areas. (Widrow, interview)

After looking at Widrow's group's attempts at solving the problem of training multilayer nets, I would like to turn now to the SRI group (the other main center of early neural network research). The SRI researchers also saw that problem as one of the main issues of early neural network research. (In Section 3, I showed that Rosenblatt was also aware of this problem.)

The SRI researchers were aware that more classification power could come from making the connections of the second layer of their Minos machine adjustable, but they could not develop adequate training techniques for that. This was openly admitted by Nils Nilsson (one of the leading members of the group) in his book *Learning Machines* (Nilsson, 1965). Nilsson himself (interview) and Charles Rosen (interview) confirmed it to me.

> 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)[11]

Of course, training multilayer systems was not the only problem faced by SRI neural network researchers. Much of what was said in the previous section about the problems of the single-layer perceptron also applied to Widrow's madaline machines and the SRI group's Minos.

5. Interpretative Flexibility

I said earlier that, as the perceptron controversy increased and developed, Minsky and Papert decided to intervene in it decisively and show once and

[11] Rosen continued as follows: "... This problem has now been solved, and in the past ten years has led to an explosive increase of interest in the theory and applications of neural nets, a field that will probably remain important in Artificial Intelligence from now on" (Rosen, interview).

for all the limits of the perceptron. Minsky and Papert's intervention was one of the most important episodes in the perceptron controversy. In this section I look in detail at some of Minsky and Papert's main arguments against the perceptron, and I show that they were interpreted rather differently by researchers in favor of and contrary to neural network research. In the following sections I will analyze the importance of Minsky and Papert's arguments in the closure of the perceptron controversy.

Minsky and Papert had worked separately on neural networks long before they embarked on their common project in the early 1960s.[12] Approximately at the time when Seymour Papert went to MIT (in 1963, when the controversy had reached its highest level), he and Marvin Minsky decided to carry out a study that would show the limitations of perceptrons beyond doubt. They decided to make a decisive "move" in the perceptron controversy, a move aiming at showing beyond doubt the limitations of the perceptron (and of neural network research in general). With this move, Minsky and Papert aimed at mobilizing as many (and as good) arguments as possible in their favor, so that their position could not be contested by their opponents.

An important part of the motivation for carrying out this study came from the fact that some neural network researchers were trying to get funding to build more powerful machines.

> 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 work. 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)

> 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. (Papert, 1988, pp. 4–5)

> 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. (Bernstein, 1981, p. 100)

[12] I discuss Minsky's early work on neural networks elsewhere (Olazaran, 1991, Section 2.1). For Papert's early involvement, see Papert (1988, p. 11).

Similar statements by Minsky and Papert can be found elsewhere.[13]

But before turning to the discussion of Minsky and Papert's study, it is *very important* to clarify some issues about the *time frame of the perceptron controversy*. The final results of Minsky and Papert's study about the limitations of perceptrons were not published until 1969 (Minsky and Papert, 1969). Nevertheless, the main points of their critical arguments against perceptrons (and neural networks in general) were well known by the mid-1960s and had critically affected the development of neural network research by then.

Minsky and Papert's (1969) study was the "final push," so to speak, for the closure of the controversy, but many of the main events in that closure—including the crises of both the SRI Minos project and Widrow's madaline projects—happened in the mid-1960s.

One reason for delaying publication of their results was that some of the mathematical work was harder than expected. Minsky and Papert delayed publication until they had given an elaborated mathematical form to many of their points.

> 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. (McCorduck, 1979, p. 89)

> 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]. (Minsky and Papert, 1988, pp. 249–250)

[13] "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 century 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" (Minsky and Papert, 1969, pp. 18–20).

> 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. (Papert, 1988, p. 11)

What I would like to emphasize here is that the main points of Minsky and Papert's arguments against the perceptron were well known by the mid-1960s, and that those arguments had had a critical effect on neural network research by then. This was confirmed to me by the researchers of the early neural network period whom I could interview (see Appendix 1). The following remark by Dreyfus and Dreyfus, therefore, seems correct.

> About 1965, Minsky and Papert, who were running a laboratory at MIT dedicated to the symbol-manipulation approach and therefore competing for support with the perceptron projects, began circulating drafts of a book attacking the idea of the perceptron. (Dreyfus and Dreyfus, 1988, p. 21)

In the rest of this section, I will analyze Minsky and Papert's main critical arguments against the perceptron as they finally appeared in their 1969 book *Perceptrons* (Minsky and Papert, 1969) because they are best expressed there, but I would ask the reader to read them as if they had been written in the mid-1960s. I think that this is justified because, even though the book had an effect of its own when it was published—it was the *last word* in the perceptron controversy, the final push for its closure—most of the arguments published there had had a very important effect on the evolution of neural network research by then.

I would like to distinguish two different issues in Minsky and Papert's arguments against the perceptron (and neural network systems in general). One is their work on the limitations of single-layer perceptrons. The other is their "intuitive judgement" (in their own words) about the possibility of developing a learning algorithm for multilayer perceptrons. In popular versions of the history of neural networks, Minsky and Papert's Perceptrons study is usually taken as the proof that neural network research had so many problems that it was not worth pursuing, and that therefore it had to be abandoned. However, in my opinion, that common view was the *result* of the closure of the perceptron controversy, and not its cause. Before then, things were not clear at all. The two main parts of Minsky and Papert's attack on the perceptron were open to interpretative flexibility (see Section 1 for comments on this concept). This was so not only in principle; neural network researchers did try to take advantage of that interpretative flexibility *in practice*. The problem was—as always is—whether neural network researcher's response to Minsky and Papert's criticism was *strong enough* (I will get to this issue in Section 8).

In my analysis of Minsky and Papert's (1969) discussion of the problems of neural networks I will use terms like *reverse salient* and *anomalous problem*. I explained what I mean by reverse salient in Section 4. Thomas Kuhn (1970) used the term *anomaly* extensively in his historical studies of science. For Kuhn, anomalies are experimental results that do not fit within the accepted categories of a scientific theory. Kuhn argued that, when anomalies pile up and are important and persistent, their solution may require severe adjustments in the theoretical and methodological apparatus of a paradigm. But this is not the sense in which I would like to use the term anomalous problem here. By anomalous problem I mean a research problem that both: (a) resists solution within a scientific approach, and (b) has an acceptable solution within a competing paradigm. Nevertheless, I maintain the view that notions like *resistance to solution* and *acceptable solution within a competing paradigm* are the product of a process of social negotiation and decision. Scientific and technical problems are often evaluated differently by the diverse groups participating in a controversy.[14]

The concept of anomalous problem will be used in the discussion of the limitations of the (single-layer) perceptron, which I would like to distinguish in Minsky and Papert's study. The concept of reverse salient will be used in the discussion of the next issue, namely Minsky and Papert's challenge (or "intuitive judgement") about more complex perceptrons.

Some of the main problems of the single-layer perceptron analyzed in-depth by Minsky and Papert—e.g., the parity and connectedness problems—can be seen as anomalous problems. But it is important to emphasize that problems like parity and connectedness became anomalous not because of some intrinsic or necessary property, but because early neural network researchers were not strong enough to resist Minsky and Papert's (1969) conclusions about their "anomalous character." Those problems became anomalous because they were important in the closure of the perceptron controversy. In other words, the anomalous character of those problems was in part the result of that closure, and not its cause. There was nothing *intrinsically* anomalous in problems like parity and connectedness. I will show later that, before the perceptron controversy was closed, there were different interpretations of those problems, some of them favoring the continuation of neural network research.

Figure 6 shows the single-layer perceptron analyzed by Minsky and Papert (1969) in their study.

[14] My definition of anomalous problem is a sociological reinterpretation of Larry Laudan's definition of anomaly (see Laudan, 1977, p. 29).

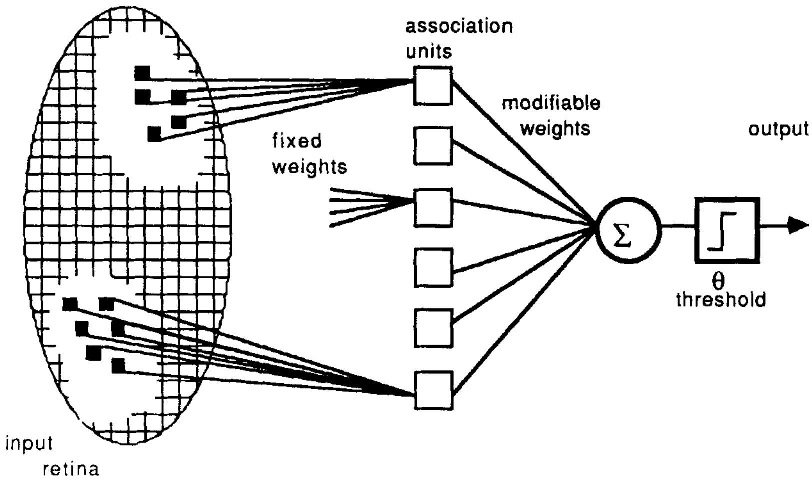

FIG. 6. Perceptron studied by Minsky and Papert.

Minsky and Papert introduced some restrictions in the perceptron they studied. A very important restriction affected the connections from the input units to the association units (the layer of fixed connections in Fig. 6). This restriction was a consequence of their definition of computing in a perceptron. They defined the computation realized by a neural network system as a parallel combination of local information. Minsky and Papert thought that, for this computation to be interesting or effective, it had to be simple in some meaningful sense.[15]

The computation realized by the output unit in Fig. 6—a sum of incoming weighted activation in parallel plus a comparison with a threshold—satisfied Minsky and Papert's criterion. In the case of the association units, Minsky and Papert's interpreted their "simple combination of local information" criterion as implying 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. (Therefore, the order of the perceptron of Fig. 6 is 6.)

[15] They 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_a which are 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" (Minsky and Papert, 1969, p. 9).

The implications of Minsky and Papert's criterion of "conjunctive localness" are better understood by looking at the parity and connectedness problems. The parity problem consists of saying whether the number on activated inputs in a perceptron retina (or set of input units) like the one in Fig. 7 is odd or even. (In Fig. 7, the number of inputs that are "on" is odd, namely 13 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. Minsky and Papert (1969, ch. 3) showed that the order required for a single-layer perceptron like the one in Fig. 6 to compute parity was the whole retina, that is at least one association unit had to receive connections from all the input units. But, they went on, if one association unit had to "look at" all the input units in the retina, then the computation realized by the perceptron was not based on a combination of *local* information, and therefore the "conjunctive localness" criterion was not satisfied by the single-layer perceptron.

The second of the single-layer perceptron's problems studied by Minsky and Papert (1969) that I would like to look at here is the "connectedness" (or figure-ground) problem. I showed earlier (Section 3) 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., whether they are connected to each other) or not. The input pattern appearing in the retina of Fig. 8 is connected (all the activated input units belong to the same object).

Minsky and Papert (1969, ch. 5) claimed that the order required for the perceptron of Fig. 6 to compute the connectedness property exceeded practical and acceptable limits too. This order grew "arbitrarily large" as the input

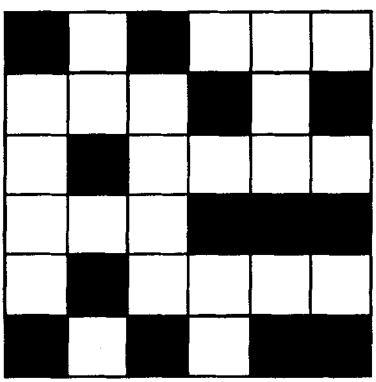

FIG. 7. Odd number of activated inputs.

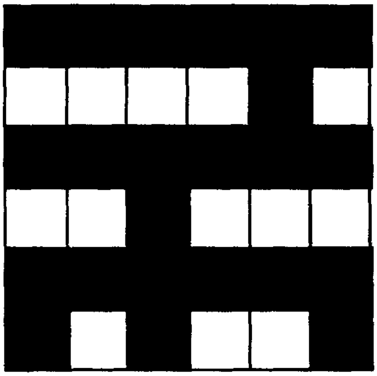

FIG. 8. Connected.

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 $\Psi_{\text{connected}}$ [the connectedness predicate] . . . *Any* perceptron for this predicate on a 100×100 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 and Papert, 1969, p. 17)

> Of course, if some φ [association unit] is allowed to look at *all* the points of R [retina] then $\Psi_{\text{connected}}$ can be computed, but this would go against any concept of the φ's as local functions. (Minsky and Papert, 1969, 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. This problem is equivalent to a conventional computer program having to be rewritten when changing the size of the task.[16] Minsky and Papert 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

[16] Aleksander and Morton made this comparison, "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" (Aleksander and Morton, 1990, p. 41).

limited area of the retina. Minsky and Papert showed that diameter-limited perceptrons could not recognize the connectedness of a figure. The simplified version of their proof is easy to visualize (see Minsky and Papert, 1969, pp. 12–14).

The anomalous (in the sense I am using this term here) character of a problem increases if researchers *agree* to compare the solution (or the lack of solution) given by a tradition of research to that problem with the solution given by a competing tradition of research. One important move in Minsky and Papert's argumentative rhetoric (1969, ch. 9) was to claim that problems such as parity or connectedness could be solved easily using conventional algorithms in serial computers.

> The predicate $\Psi_{\text{connected}}$ 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 and 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. (Minsky and Papert, 1969, p. 227)

By emphasizing that parity and connectedness could be easily computed by conventional algorithms in serial von Neumann computers, Minsky and Papert were trying to mobilize two very powerful factors in favor of their argumentative position: the symbol-processing approach to AI and the digital computer.[17]

But the importance of problems like parity and connectedness was not so clear for neural network researchers. The interpretative flexibility of the problems of parity and connectedness can be better shown with an example. Consider Figs. 9 and 10. It is not immediately obvious whether this type of

[17] Aleksander and Morton described some simple algorithms for computing the parity and connectedness of the retinas appearing in Figs. 7 and 8: "(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" (Aleksander and Morton, 1990, pp. 39–40).

FIG. 9.

figure is connected or not. Figure 9 is not connected; Fig. 10 is. Consider now the white background as a figure, and look at the center of the drawing (this is another possible example). The object appearing in Fig. 9 is now connected; the one in Fig. 10 is now unconnected. But this is not obvious the first time one looks at those objects. A conscious, "serial" process is necessary to determine the connectedness of these figures.

Early neural network researchers conceded that perceptrons were not very good at recognizing parity or connectedness, but—they went on—neither are human beings. They claimed that, far from being a problem, this could be a positive characteristic of perceptrons. For an example of the difficulty of the recognition of parity for humans, see Fig. 11. Are the number of little

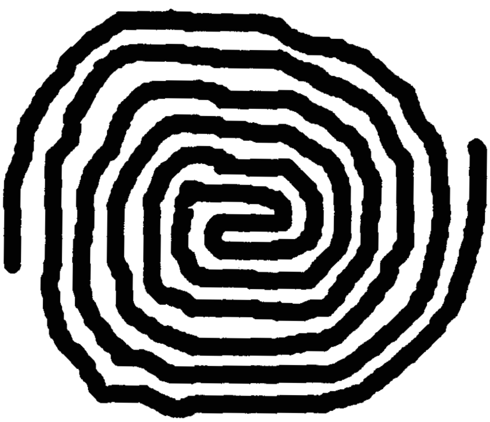

FIG. 10.

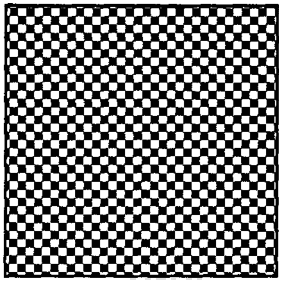

FIG. 11

black squares on it odd or even? It is impossible to say without carefully, slowly, and "serially" counting them.

The relative importance of the problems of parity and connectedness (and their possible "anomalous" character) was open to interpretative flexibility. Perceptron researchers 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 worrying (let alone anomalous) after all, because human beings are not good at recognizing parity or connectedness either. David Block, a mathematician from Cornell University who was a colleague of Rosenblatt in the Perceptron Project, replied 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. (Block, 1970, p. 517)

The relationship between human information processing (at the psychological or neurobiological level) and machine information processing has been a constant rhetorical resource throughout the history of AI. If the machine solves a certain task well, then one does not care about the neurobiological or psychological plausibility of the system's architecture and functioning. 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.

Thus, perceptron researchers replied to Minsky and Papert's rhetoric with their own rhetoric. The point was, of course, whether this argumentative rhetoric was strong enough to contest Minsky and Papert's criticism. Minsky and Papert focused their attention on problems (such as parity and connectedness) that were favorable to their position. In his reply, Block focused on a set of different arguments favoring the neural network position. Thus, the importance of the problems of parity and connectedness was open to interpretative flexibility. Different groups of scientists, with different goals and interests, interpreted them in diverse ways. For Minsky and Papert, they were very worrying, anomalous problems. However, neural network researchers not only claimed that they were not very important, but also that they were in a sense "successful evidence" in favor of their own approach.

I showed earlier that neural network researchers were aware of the difficulties that single-layer perceptrons had in computing predicates like exclusive-or and connectedness well before Minsky and Papert started to circulate their drafts in the mid-1960s. But for them the existence of these problems was not a strong argument against the neural network approach. In their view, single-layer perceptrons were only the first stage of neural network research. Rosenblatt had openly admitted the limitations of single-layer perceptrons. In particular, his insistence on the connectedness problem was almost repetitive (see Section 3). But Rosenblatt and the other early neural network researchers had an approach to the limitations of the perceptron that 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 carrying out further 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 emphasized the positive properties of the single-layer perceptron (e.g., its learning algorithm, its brain-like character, its distributed memory, its resistance to damage, its parallelism), and claimed that further research on more complex models was needed in order to realize tasks more complex than those that could be carried out by the perceptron.

> ... 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. (Block, 1970, pp. 513–514)

Here Block emphasizes the "promising future" of neural network research. The opinions of other neural network researchers of the time were similar in certain respects. Thus, Widrow complained that Minsky and Papert had defined perceptrons so narrowly that they could prove that those neural network machines could do nothing. He also emphasized that he and his colleagues were working on networks much more complex than the single-layer perceptron.

> 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)

Thus, one of the main arguments used by neural network researchers was that they were working on neural network systems more complex than those studied in detail by Minsky and Papert. It is to this issue (the second one of the two I have distinguished in Minsky and Papert's critical study) that I would like to turn to now.

On the issue of perceptrons more complex than the single-layer one, Minsky and Papert made only a few comments in their 1969 book. They formulated a (by now famous) pessimistic "intuitive judgement" about the possibility of developing efficient techniques for training multilayer systems.

> 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 and Papert, 1969, pp. 231–232)

I showed in earlier sections that the problem of learning in multilayer neural networks had been on neural network researchers' agenda long before Minsky and Papert embarked on their criticism of neural networks. In fact, it was one of the main problems of neural network research from the late 1950s onward. Nevertheless, because of the important role played by Minsky and Papert's critical arguments in the closure of the perceptron controversy, the short comments they made there about the problem of learning in multilayer networks emphasized its importance even more. By focusing only on that aspect of multilayer systems, they helped "construct" (so to speak) its "reverse salient" character. (I will come to this issue in later sections.)

From the point of view of this section, I would like to stress that Minsky and Papert's short comments on the problem of learning in multilayer networks were indeed open to interpretative flexibility. If a short intuitive judgement is not open to interpretative flexibility, then nothing is. Therefore, the question now is to analyze the process of closure of the controversy, i.e., the process through which the interpretative flexibility of Minsky and Papert's results and critical arguments about the problems of both single-layer and multilayer perceptrons was reduced and the controversy closed. I will discuss this next.

6. Closure of the Controversy 1: Widrow's Group

The crisis of early neural network research deepened as acceptable solutions to some of the main technical problems of the field could not be easily found and criticism by researchers in favor of the symbol-processing approach increased. In this section I will look at the crisis of Bernard Widrow's neural network research group at Stanford University. I will later look at other research groups.

The crisis of Widrow and colleagues' neural network research project reached its peak toward 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 neural network ideas (mainly the adaline and the LMS algorithm) in areas like adaptive filtering and adaptive signal processing started to be more successful. One first successful application of Widrow's adaline and LMS neural network techniques outside neural networks was in adaptive antennas.[18]

> 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. We were finding good applications for our

[18] See Widrow *et al.* (1967).

> neural network devices. 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 it, it 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 quantizers [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. 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. In the second half of the 1960s, R. W. Lucky and his team at Bell Laboratories applied adaptive filters to telephone systems in areas such as adaptive equalization in high-speed modems, and echo canceling in long-distance telephone and satellite circuits.[19]

> 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 equalization. And they must adapt to that particular phone line, because every phone line is different. 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 Institute of Electrical and Electronic Engineers' (IEEE) Alexander Graham Bell Prize in 1986 for his "exceptional contributions to the advancement of telecommunications."

What is most interesting from the point of view of this chapter is that, after the crisis of early neural network research, both the LMS algorithm and the adaline—first developed in the context of neural network research—

[19] See Lucky (1965), Lucky *et al.* (1968), Widrow and Lehr (1990, pp. 1415–1416). 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 that has an adaptive filter has to be installed in the receiving telephone. On the other hand, by using an echo canceller with an adaptive filter capable of performing LMS weight adjustment, telephone line echo is canceled, and it is 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 especially useful for long-distance and satellite telephone communication because of the longer time delays produced in these cases. It can also be installed in modems, so that echo can be avoided in high-speed digital data transmission by telephone as well.

were rather successful in other engineering areas, leading, as Widrow and Lehr (1990, pp. 14–16) put it, to "major commercial applications."

> I didn't invent the echo canceller, I didn't invent the adaptive equalizer. Lucky invented the adaptive equalizer. 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 the adaptive equalizer 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 equalizer. 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. (Widrow, interview)

In Section 12, I will show the amazing way in which Bernard Widrow was "rediscovered" in the recent re-emergence of the neural network field. From the point of view of this section, it is interesting to note that, because of the crisis of neural network research, Widrow and his colleagues decided to look for applications of their neural network devices and techniques outside the neural network field. Techniques like the adaline and the LMS learning algorithm, which had been first developed in neural network research, became parts of new configurations of technology in a different area of applications, namely telecommunications. Therefore, they did not disappear after Widrow abandoned neural network research.

7. Closure of the Controversy 2: The SRI Group

I would like to turn now to another main neural network center of the time, namely Stanford Research Institute (SRI). By the mid-1960s the crisis of neural networks was affecting this group strongly. Financial support for their neural network project was running out, and the limitations of Minos—their single-layer neural network machine, probably the biggest one from the early neural network period—were becoming increasingly apparent. I showed earlier that the SRI researchers were well aware of the problems of single-layer and multilayer neural networks. Problems like that of training multilayer systems could not be easily solved, and criticism from researchers favoring the symbol-processing approach was increasing. On the other hand, symbol-processing AI (the long-time rival of neural networks) was emerging with considerable success by then (for a study of the emergence of symbolic AI, see Fleck [1978, 1982]). In this context of crisis, the SRI researchers decided to abandon neural networks and to start a robotics project within the symbolic perspective. They had a feeling that they had gone as far as they could within the neural network paradigm.

> 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).

> 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 group started to work on robotics and machine vision within symbol-processing AI. Machine vision was a part of symbolic AI, but it had its own peculiarities. It was closer to SRI researchers' interests in perception and vision than symbolic AI models of "higher" cognitive processes. Work on machine vision within symbolic AI, mainly that of Larry Roberts (1963) of MIT's Lincoln Laboratory, had a significant impact on the SRI researchers, especially on those more oriented toward pattern recognition, like Richard Duda and Peter Hart. Within symbolic AI, machine vision emphasis was on scene analysis of digitized pictures in terms of lines, edges, vertices, relative brightness, and linguistic descriptions, rather than on neural network-like pattern classification. For scene analysis, the computer system needed an internal representation or model of its surrounding block world (i.e., its knowledge and expectations about that world). The SRI researchers were impressed by this type of AI research on machine vision.

The trends toward the symbolic approach in machine vision affected other researchers too. One significant case in this respect is David Marr's. In the late 1960s Marr had carried out some research related to neural networks (Marr, 1969, 1970, 1971), but in the early 1970s he changed his direction of research toward the symbolic paradigm. His words reflect 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, 18–19)

Robotics was also closer to the SRI group's interests than other areas of symbolic AI. One of the most interesting aspects of these robotics projects of the late 1960s and early 1970s was the integration in the same system of interacting subsystems for vision, planning, and object manipulation.

The criticisms made by researchers who opposed neural network research in the perceptron controversy had a big influence in SRI researchers' decision of changing their research direction from neural networks to symbol-processing AI. In particular, Minsky's active criticisms seem to have been quite important in this respect.

> 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)

Minsky had been aware of the neural network activity at SRI.

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

For Minsky, the change of research direction at SRI in the mid-1960s shows that by that time neural network research "was really dead":

> 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)

But the SRI group's transition from neural networks to symbolic AI was not something that could be done immediately. It is interesting to point out in this respect that the SRI researchers had to hire Bertram Raphael (a former student of Minsky) to teach them LISP (list processing language) and help them in their robot project.[20] LISP had been developed by John McCarthy in 1960, and had already become the most widely used programming language in symbolic AI. The fact that the SRI neural network

[20] "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). "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" (McCorduck, 1979, p. 231).

researchers had not used LISP before they switched to symbolic AI shows the distance that separated neural networks and symbolic AI in the early 1960s.

The difficulties that Rosen found in getting funding for their new project from DARPA are another interesting aspect of the SRI group's transition from neural networks to symbolic AI. Rosen had to "sell" their robot project to people who did not like neural networks, that is to people who had been convinced by the anti-neural network position in the perceptron controversy.

> 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 [symbolic] artificial intelligence. It took about three to four 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 program 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 six or seven years we built, I'd say, the first really smart robot in the world. This robot ("Shakey"), served as a test bed for studies and experimental verification in important sub-fields in AI, such as pattern recognition, scene analysis, natural language processing, navigation and obstacle avoidance, problem solving, and more. (Rosen, interview)

There was some controversy about how funding for the SRI robot project was given and (years later) cut, but looking at it is out of the scope of this chapter.[21] But that was not the only controversy in which that project was involved. The SRI Shakey robot became quite popular outside the scientific community thanks to a polemical article published in *Life* magazine (Darrach, 1970). This article, which was widely read (McCorduck, 1979, p. 235), contained some rather optimistic claims about artificial intelligence

[21] "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 McCorduck (1979, pp. 233–235). The cut in robotics funding was a wider 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.

(this is certainly another episode of allegedly "exaggerated claims" within AI—but many of the considerations I made in Section 2 about Rosenblatt's exaggerated claims apply here too).[22]

In their Shakey robot, the SRI researchers combined perceptual, motor-control, problem solving, knowledge-representation (including internal representations and models of the block world), and plans (McCorduck, 1979, pp. 223–235).[23] Some robotics projects of that period (e.g., those at MIT, Stanford University, and Edinburgh University) were "hand-eye" systems, but Shakey was a mobile robot that 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 (for details of Shakey see Nilsson and Raphael [1967] and Raphael [1976, pp. 275–282]).

The change in SRI neural network group's direction of research reflects clearly the situation of deepening crisis of neural networks and the (by then quite successful) process of institutionalization of symbolic AI. For the SRI group, the Shakey project meant the change from neural network research

[22] Darrach's (1970) 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, 66). Some of the SRI researchers criticized this article heavily, and Minsky denied the quotations attributed to him (McCorduck, 1979, pp. 234–235). However, it seems that Darrach's article was beneficial for the SRI researchers, at least in terms of getting funding. C. Rosen 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).

[23] ". . . '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).

to symbolic AI, and in particular to machine vision and robotics. Symbolic representation and knowledge issues became increasingly important for them, and they devoted their attention to combining problem solving and reasoning mechanisms with sensory and vision processes. It is also interesting to point out that, with their robot project, the SRI group became one of the leading groups in symbolic AI.

> 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 steam. (Hart, interview)

Thus, the effects that the crisis of the neural network had on the SRI group were different from the effects it had on Widrow's group. Unlike Widrow, the SRI researchers chose to stay within AI. But staying within the AI field meant switching to the symbol-processing paradigm *and* abandoning neural networks. The emergence and institutionalization of symbolic AI in the early and mid-1960s was an important "closing factor" in the neural network controversy. From the point of view of this chapter, it is important to emphasize that, in the early stages of the history of AI, neural networks and symbol-processing were seen by most as alternative, excluding approaches. Therefore, embracing the emergent symbolic AI approach meant abandoning neural networks completely. I will come to this issue in the next section.

8. Closure of the Controversy 3: Rosenblatt

After discussing how the crisis of early neural network research affected both Widrow's group at Stanford University and Rosen's group at SRI, I would like to turn now to Rosenblatt's response to Minsky and Papert's critical challenge. Unlike Widrow and his colleagues or the SRI researchers, Rosenblatt did not abandon neural networks in the mid-1960s. He and his colleagues tried to exploit in their own favor the interpretative flexibility of Minsky and Papert's criticism of neural networks, but they were increasingly isolated, and could not stop the process of closure of the controversy. In this section, I also discuss some factors that were very important in the closure of the perceptron controversy, such as the emergence of symbolic AI and developments in computer technology.

I pointed out in Section 1 that, *in principle*, scientific controversies could always go on. The point is then to analyze how the plausibility of the positions involved in a controversy evolve, and how the possibility (in principle) of going on arguing is *in practice* reduced and controversies are closed. This process of closure of controversies is especially important in the generation and validation of scientific knowledge. In this section I study the process through which the interpretative flexibility of the criticisms of the perceptron was closed against the position maintained by (a decreasing number of) neural network researchers.

One important issue here is that, in part as a result of the perceptron controversy and the criticisms of neural networks by researchers favoring symbolic AI, Rosenblatt was unable to get economic support for his projects from the funding agencies in the mid-1960s. I showed earlier that the rise of the perceptron controversy goes back to ONR's funding of Rosenblatt's project in the late 1950s. It is very important to remember here that the issue of funding was one of Minsky and Papert's main motives for starting their critical perceptrons project. (I looked at this in the beginning of Section 5.)

Information about funding for the perceptron project throughout the 1960s is not easy to obtain. Nevertheless, I was able to talk with two people who were at ONR at the time: Marshall Yovits and Marvin Denicoff.[24] Yovits was at ONR in the late 1950s and early 1960s, and was responsible for the funding for Rosenblatt's projects during that period. Denicoff was at ONR from the mid-1950s to the early 1980s. Their testimony is important, because ONR was probably the only source of financial support for Rosenblatt's projects over the years.

Rosenblatt's relationships with the U.S. military do not seem to have been too easy. One factor that might have had some kind of influence on this issue was Rosenblatt's involvement in the peace movement at the time of the Korean War. As a consequence of that, he never got security clearance from the U.S. military.[25]

[24] Marshall Yovits is now at Purdue University, Department of Computer and Information Science, Indianapolis, Indiana. Marvin Denicoff now works at Thinking Machines Corporation, Chevy Chase, Maryland.

[25] "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 favor 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, interview).

Yovits was at ONR in the late 1950s and early 1960s, when interest in the perceptron was at its highest level.[26] After that Rosenblatt was not as successful in getting funding from the U.S. military agencies. It seems that later in the 1960s he was not able to get funding for a "large" perceptron research project. Financial support for Rosenblatt's perceptron projects was cut at some point, although it is difficult to establish the exact date at which this happened. Nevertheless, it seems to be the case (many of my interviewees—see Appendix 1—confirmed this) that in the second half of the 1960s, Rosenblatt was unable to get economic support for his projects from ONR and DARPA (then ARPA).

> 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 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 this quotation, Yovits seems to be implying that Rosenblatt was not successful in convincing another funding agency ("our programmes were small"). That agency could well have been DARPA. "Big dollars" for AI-like research in the 1960s (and thereafter) came from DARPA. 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 DARPA (therefore, he was well informed about DARPA'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

[26] 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-organization," which were held in 1959, 1960, and 1962. The proceedings were published as (Yovits and Cameron 1960), (von Foerster and Zopf, 1962), and (Yovits *et al.* 1962), respectively. Yovits himself was one of the organizers of these conferences, and contributions presented there are a good sample of the work that was being carried out in neural networks and related fields in the late 1950s and in the first years of the 1960s. They also show the variety of approaches to the brain-machine problem, which were tried within the cybernetics movement.

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 program, 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's remark about Rosenblatt not offering a big pay-off in terms of applications could be related to the general patterns of science policy in the United States in the 1960s. From the late 1950s to the mid-1960s, funding for science in the U.S. 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.[27] ONR's support for Rosenblatt's perceptron in the late 1950s and early 1960s happened within this context. At that time ONR had a reputation of working without worrying too much about the pay-offs in terms of applications.[28] The situation did not last long, however, and the growth rate in funding for scientific research decreased significantly from

[27] The opening of the space race with the launching of the Sputnik satellite in October 1957 by the Soviet Union was the catalyzer. Support for science increased to unprecedented levels. ARPA itself was created within the Defense Reorganization Act of 1958, a reaction to the Sputnik launch.

[28] "[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).

about 1965.[29] The growing concern in the Defense agencies with the short-term 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" (Dickson, 1988, p. 30).

This context of science policy did not favor Rosenblatt's chances of being funded. In the "Tribute to Frank Rosenblatt," celebrated in July 1971 after Rosenblatt's death, Richard O'Brien, head of the Division of Biological Sciences of Cornell University, Ithaca, New York, (where Rosenblatt was working at the time), made the following comment in his speech:

> ... It was only a few years ago that he [Rosenblatt] 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. (Congressional Record, 1971, p. 3)

In the mid- and late 1960s, Rosenblatt worked on perceptrons and some other projects.[30] From the point of view of this chapter, it is important to emphasize that, unlike Widrow's group or the SRI group, Rosenblatt did not abandon neural network research. Despite the crisis of neural networks and the process of closure of the controversy in the mid- and late 1960s, Rosenblatt kept working on perceptrons until his final years.

> ... 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 synthetic 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)

[29] "... 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" (Dickson, 1988, pp. 5–6, 123).

[30] He worked on 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).

The importance of Rosenblatt as an individual cannot be forgotten in an analysis of the closure of the perceptron controversy. I stated in earlier sections that Rosenblatt was often the leader of the neural network position in the 1950s and 1960s. His death in a tragic boating accident in 1971 left neural network research without its most enthusiastic and charismatic defender.[31] Rosenblatt's death makes research on the reaction of Rosenblatt's group to the crisis of early neural networks (and to Minsky and Papert's criticism of the perceptron) much more difficult. Some elements of that reaction can be inferred from David Block's review of Minsky and Papert's 1969 *Perceptrons* (Block, 1970) (some aspects of this response were discussed in Section 5).

Block's response to Minsky and Papert was based on two main points. First, Block accused Minsky and Papert of trying to control the focus of the debate for their own advantage. He criticized 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). He also accused them of taking "a temporary handhold" (i.e., the single-layer perceptron) as if it were the "final" neural network structure (Block, 1970, p. 514). Furthermore, Block attacked at the point where Minsky and Papert's argument was most open to interpretative flexibility, namely their comments on the problem of learning in multilayer perceptrons. Block was optimistic about the promising possibilities of more complex perceptrons "with several layers, feedback, and cross coupling" (Block, 1970, p. 513). He tried to gather as many argumentative elements as he could in favor of the neural network position. The number of authors cited in the following quotation is an example of this. Block's reply to Minsky and Papert's pessimistic "intuitive judgement" on multilayer perceptrons was an appeal to the promising sides of neural network 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 *et al.*, [1962], Konheim . . .). Even more suggestive are the multilayer machines with feedback (the *C*-systems and *F*-systems of Rosenblatt [1967]). The models studied extensively by Grossberg, . . . 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 and Papert, 1969]. The same can be said of other brain models, such as those of Kabrisky . . . or Baron. . . . From this point of view, the potential capabilities of Perceptrons are still mostly unexplored. (Block, 1970, pp. 516–517)

[31] 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*, 1971).

But Block's efforts were not enough. His arguments—and in general the arguments used by the researchers in favor of neural networks—did not have enough credibility. Closure against their position was approaching. There was nothing *intrinsically* superior in Minsky and Papert's pessimistic "intuitive judgement" about multilayer perceptrons as compared with Block's conclusion that "the potential capabilities of Perceptrons were still mostly unexplored." However, Rosenblatt and his colleagues were increasingly isolated, and very much on the losing side of the controversy.

The key to the process of closure of the perceptron controversy was, in my view, the linkage between the criticism of neural networks and certain "closure factors," the most important of them being the emergence of symbolic AI. When an area of research is emerging, it has a special need of resources and legitimation, both inside and outside the research community. In this sense, the closure of the perceptron controversy was more important than what is usually thought for the development of symbolic AI. In the late 1950s and early 1960s, neural networks and symbolic AI were both in their early phase (they both originated in the post-war cybernetics movement). In such a situation, when two paradigms are emerging and have a low level of development, competition for resources and legitimation can be particularly strong. This may well have been the reason with symbolic AI researchers were especially interested in criticizing neural network research, and in closing the neural network controversy once and for all. The view resulting from that closure—that there was no credible alternative to symbol-processing—helped legitimize the institutionalization of the symbolic approach.

In the context of competition for resources and legitimation between neural networks and symbolic AI in the 1960s, Minsky and Papert's criticism of neural networks was interpreted as the "final proof" showing the lack of validity of neural networks and the adequacy of the remaining approach (symbol-processing) as the *only* approach to AI. The following quote from Allen Newell and Herbert Simon's important "Computer Science as Empirical Enquiry: Symbols and Search" paper on the foundations of symbol-processing AI is an example of this view of symbol-processing as the "only approach" to AI:

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

I think that the closure of the perceptron controversy was the origin of this view. That closure is the "marker event" that Newell was looking for in his paper on the history of AI (Newell, 1983). There Newell indicated that the process of emergence of symbolic AI was "essentially complete by 1965," although he could not find a "marker event." This marker event was, in my view, the crisis of neural network research, which happened in the mid-1960s, and the closure of the perceptron controversy, which started then and was finally completed by the end of the 1960s.

> 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***. . . . [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. (Newell, 1983, pp. 201–202, emphasis and bold added)

Thus, the crisis of neural networks and the closure of the perceptron controversy may well have been more important than what most historical accounts of AI usually concede. As a consequence of that closure, the acceptance of the symbol-processing approach was linked to the rejection of neural networks. Both Minsky and Papert have admitted that a link was made (by many researchers and people in the funding agencies) between their *Perceptrons* book and the rejection of the *whole* neural network paradigm (including multilayer systems). Papert's words about "universalistic attitudes" (see quotation to follow) in the AI culture are especially interesting in this respect, and Minsky is reported to have "regretted the chilling effect of his Perceptrons book on neural networks" in a neural network conference in 1988 (*Alternative Computers*, 1989, p. 51).

> 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, 1988, pp. 7–8)

Of course both the universalistic AI culture mentioned by Papert and the interpretation of Minsky and Papert's (1969) study as showing that *the whole* neural network approach had to be abandoned were social phenomena. These phenomena went well beyond Minsky and Papert's involvement in the controversy as individuals.

So far, I have looked at one closing factor of the perceptron controversy (perhaps the most important one), namely the emergence of symbolic AI. I would like to turn now to another very important closing factor, i.e., the "linkage" between symbolic AI and digital computer technology. Four things could be said about this linkage. First, the von Neumann style of computation is based on the same information processing principles as those of the symbolic-processing approach.[32] Second, from the beginning the digital computer was the experimentation tool of symbolic AI researchers (a great part of research activity in symbolic AI consisted of digital computer simulations using list processing programming languages like LISP). Third, in the late 1950s and early 1960s, symbolic AI researchers at DARPA-funded centres like MIT, Carnegie-Mellon University, and Stanford University had a privileged access to digital computer resources (see Fleck, 1982). Finally, the von Neumann computer was somehow "emblematic" for symbol-processing AI researchers, because it showed that one could study intelligence and still be a "materialist." Computer mechanisms (hardware and software) were capable of carrying out certain tasks for which intelligence was thought necessary.

The impressive developments in digital computer technology from the mid-1960s onward strengthened the position of the symbol-processing approach. Developments in digital computer technology from the late 1960s onward have been spectacular (for a review, see Molina, 1987, ch. 2). Hardware developments include miniaturization 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).

Analog and neural network technologies were very much on the losing side of the computer technology race. The eclipse of analog computers in

[32] Although, of course, symbol-processing is compatible with some degree of parallelism, and therefore it does not have to be serial, as in the von Neumann computer.

the mid-1960s is especially interesting here. The history of the analog computer goes back (at least) to the 1930s, but its "golden age" was the 1950s and early 1960s (*Alternative Computers*, 1989, p. 26). Analog computers were used for solving differential equations. It is interesting to note that the demise of analog computers happened approximately the same time as the crisis of neural network research. The authors of *Alternative Computers* (1989) point this out at both voltage precision problems in analog computers and the accuracy and storage capacity of digital computers, and conclude 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. (*Alternative Computers*, 1989, p. 27)

Early neural network computers, such as the perceptron, Madaline, and Minos were in a sense on the analog side (remember their modifiable weights), and neural network technology was also on the losing side with the advent and development of digital computer technology. The issue of the analog weights of Rosenblatt's perceptron is a good example of the limitations of early neural network technology. The analog weights (continuously variable quantities) of the association-to-response connections of the perceptron were implemented using motor-driven potentiometers. These devices had an important limitation: their size. The 512-weight Mark 1 perceptron occupied a whole laboratory room. 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 (these researchers developed their own solution for implementing the modifiable weights, namely magnetic cores). At Stanford University, Widrow and his colleagues developed another technological solution, the so-called "memistors." But all these analog neural network devices had important limitations.

Early neural network researchers started to use digital computers for simulating neural networks, but the association between their theoretical approach and the digital computer was much weaker than the one between symbolic AI and the digital computer. The sequential character of the von Neumann computer did not favor radically parallel approaches to AI and cognitive modeling such as neural networks. Furthermore, at that time simulating neural networks in a sequential computer did not seem to many to be the most adequate way of using the digital computing resources available. After all, Rosenblatt had characterized the philosophy of the neural network approach in radical opposition to the digital computer. Rosenblatt's comments on the issue of ("brain-style") distributed memory (where each processing unit participates in many representations and each representation is formed by the activation of many units) are a good example of this. The distributed memory of the perceptron was very different from the typical von Neumann computer memory (where data are stored in discrete, unrelated locations).

> 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 [Rosenblatt's view] that the brain operates by non-algorithmic methods, bearing little resemblance to the familiar rules of logic and mathematics which are built into digital devices. (Rosenblatt, 1962a, p. 10)

> 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)

In this section, I have looked at several factors that made Rosenblatt and his colleagues' efforts to stop the closure of the perceptron controversy vain. First, the emergence of symbolic AI, which was linked to the rejection of the whole neural network approach. Second, the funding agencies lost their confidence on neural network research and cut the funding for Rosenblatt's projects. Third, Rosenblatt was increasingly isolated in his defense of the neural network approach (other neural network researchers abandoned the field, as I showed earlier). Finally, the development of digital computer technology favored the symbolic AI approach.

Early neural network research had important technical problems, like learning in multilayer systems or the lack of computer power for carrying out large simulations. Nevertheless, these technical problems cannot be discussed out of their broader context. The same technical problems, or the same technical solutions, can be evaluated differently in different situations. Furthermore, they can be evaluated differently in the same situation by different groups of scientists. Thus, the importance of "pure" technical problems has to be put in context. I would like to finish this section with a short "science fiction" comment about the importance of one of the main technical problems of early neural network research, namely that of training multilayer networks. I think that this comment shows the importance of looking at the context when "technical problems" are analyzed.

The science fiction question I would like to ask is the following: What would have happened if someone had developed a learning technique for multilayer networks like back-propagation (which was so successful in the late 1980s) in the early period of neural network research? I agree partly with the reply that Papert gave me when I asked him this question:

> 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 *et al.*, 1986a) 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. The influential recent demonstrations of new networks all run on small computers and could have been done in 1970 with ease. . . . (Papert, interview)

I agree with Papert in that it seems that someone wanting to develop back-propagation in the 1970s would not have gotten much funding. But, although funding is very important, it is not the whole issue here. One should talk about acceptance and credibility also. Someone wanting to develop something like back-propagation in the 1970s (that is after the perceptron controversy was closed) would not have gotten much recognition or acceptance from the research community. Developing something like back-propagation would not have been enough to reopen the controversy in the 1970s. In Section 10, I will show that something like this did actually happen in the 1970s.

9. The 1980s: A Changing Context

One important effect of the closure of a controversy is the consolidation of the "balance of power" emerging from it. As the "winning" view gets institutionalized and researchers carry out their work within the accepted framework of research activity, it becomes increasingly difficult for the "losing" side to counterbalance that framework. Because of the inertia of institutions and patterns of activity (i.e., their tendency to reproduce themselves), time runs against the losing position.

> . . . 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. (Harvey, 1981, 126)

As a consequence of this inertia, the conventional character of the process of generation and validation of scientific knowledge is quickly buried and forgotten. Something like this happened at the end of the perceptron controversy. Minsky and Papert's arguments about the problems of single-layer perceptrons were widely interpreted as showing that the *whole* neural network approach had to be rejected, and symbolic AI emerged as *the* approach both to building intelligent machines and to studying and modeling cognition computationally. Some individuals continued working on neural networks throughout the 1970s, and the work they carried out was very important indeed, but they were working as isolated *individuals*; they were not powerful enough to develop a *position*.

After the closure of the perceptron controversy, activity in neural networks decreased to its minimum levels. Throughout the 1970s a small number of researchers continued research in neural networks and related topics, but they were far from the main centers of AI and cognitive science research activity, where symbol processing continued to be the dominant approach over the years.[33] Analyzing the development of neural network-like research in the 1970s is of course out of the scope of this chapter.[34] Here I would just like to point out that the main feature of this period was the retreat of neural network researchers from AI to more neuroscience- and psychology-oriented research areas.

Toward the early 1980s, several new developments started to alter this situation. Symbolic AI went from a stage of emergence and institutionalization to one of greater growth and even commercialization (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 10-year, $850 million computer technology project, in which they emphasized the importance of AI (especially natural language and knowledge-based information processing). The U.S. and U.K. governments reacted quickly by launching their own computer technology programs (namely Microelectronics and Computer Technology Corporation, DARPA's Strategic Computing program, and the Strategic Defense Initiative in the United States, and the Alvey program in Britain). This climate favored AI research.

The early 1980s were the time of the most important commercialization of symbolic AI so far, i.e., expert systems. Basically, expert systems are

[33] This was much more so in the United States than in Europe, where activity related to neural networks (but more oriented toward neurobiology and psychology) remained relatively strong. The larger relative 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 U.K. Science Research Council.

[34] Areas in which neural network researchers worked in this period include unsupervised neural networks and associative memory. Many of the most important neural network contributions of the 1970s were reprinted by Anderson and Rosenfeld (1988).

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 is laid on symbolic representation and on the ability of the computer to carry out structure-sensitive transformations of those representations. Expert systems have been (and are being) 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 commonsense and heterogeneous reasoning. In the early 1980s some researchers started to claim that neural network systems could offer solutions to some problems that were difficult to handle within the symbol-processing framework.

Computer technology developments were another important aspect of the early 1980s. Hardware developments included miniaturization, increases in computing power, and reduction in costs. The early 1980s was the time of very large scale integration (VLSI), i.e., the development of single chips with hundreds of thousands of components on them. At the same time, the limitations of the von Neumann computer architecture were becoming increasingly apparent. 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.[35] The cost of microprocessors had decreased significantly by then, but parallelism involves many issues that have not been completely understood or developed yet, like the software issue. 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, distributed AI, computer networking, parallelism in machine vision systems, and neural networks.

These trends toward parallel computing in the 1980s favored the resurgence of interest in neural networks. However, neural network research is a radical approach to parallelism, and it has to compete with other, less radical views. In neural networks, computation is defined at the subsymbolic level,

[35] 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 and Jesshope (1988, pp. 2–53).

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, researchers started to argue that the information-processing power of the brain comes from its parallelism. Feldman and Ballard (1982) formulated the so-called "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 recognizing a human face.[36] With the advent of parallel computers and supercomputers, researchers started to compare computer power and brain-style information processing.[37] But although computing power in abstract is important for neural network research (it allows increasingly more powerful simulations of and experiments with neural network systems), many problems remain (e.g., architecture and organization of the network, and learning).

I said earlier that the development of symbolic AI was linked to the development of the von Neumann computer, and that in the 1960s this association did not favor neural networks, which are naturally parallel. Recent developments in parallel computing are more favorable for neural computing. The connection between trends in computer technology and neural network research is stronger this time round, but one should not forget that neural network researchers have to compete with other (less radical) approaches to parallelism.

One subarea of symbolic AI where parallelism has sometimes been used is machine vision. In the 1970s machine vision was very much within the symbolic AI umbrella. One example of this was the change at SRI from neural networks to scene analysis in the mid-1960s (see Section 7). Another

[36] "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 . . . 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 and Ballard, 1982, pp. 484, 487).

[37] Speculations were made about the number of operations per second in the brain as compared with the most powerful computers. Sejnowski 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 to 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" (Sejnowski, 1987, pp. 206–207).

one was the change in David Marr's direction of research in the beginning of the 1970s from neural network-like research to machine vision within symbolic AI (I also mentioned Marr's case briefly in Section 7). These changes were related to the closure of the perceptron controversy. But in the early 1980s, researchers started to use some parallelism in vision research. Notions such as the parallel interaction between many local features in the interpretation of an image were an area where symbolic AI was closer than usual to neural network research (Ballard *et al.*, 1983). Hinton and his colleagues were motivated by this type of problem when they developed their "Boltzmann machine" multilayer neural network learning algorithm in the mid-1980s (Ackley *et al.*, 1985). In the changing context of the 1980s, some vision researchers started to criticize the view that the representational and abstract computational theory levels are independent of the implementation level (I will discuss some aspects of this question in Section 13).[38]

In this section, I reviewed some factors and forces that could be aligned in order to bring neural network research back to the AI front in the early 1980s. It is interesting to note that, even though neural network researchers who had continued in the field throughout the 1970s had made some very important contributions, the intervention of a new generation of researchers was decisive in bringing neural networks back to the AI arena. The Parallel Distributed Processing (PDP) group from the University of California, San Diego, was especially important in this respect. Researchers from physics (like T. Sejnowski and P. Smolensky), mathematics (like R. Williams), symbolic AI and cognitive science (like D. Rumelhart and J. McClelland), and other disciplines, saw a potential in neural networks and started to work in the field. They worked with some people who had been in the field from an earlier period (like G. Hinton), although they did not have as good communication with other such researchers (notably S. Grossberg).

10. History of Back-Propagation

The most successful technique developed in the period of re-emergence of connectionism in the mid-1980s was Rumelhart and colleagues' back-propagation learning algorithm for multilayer networks (Rumelhart *et al.*, 1986). I will look at the importance of back-propagation in the coming

[38] Vision researcher Harry Barrow (1989, p. 12) admitted that the separation between the implementational level and the other levels has sometimes been 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 [abstract computational theory level] and should affect assumptions and decisions there."

section, but before that I would like to make a short historical digression about the historical antecedents of that technique.

I said earlier that, in a way, Minsky and Papert's critical study of Rosenblatt's perceptron was a positive contribution to neural network research. Studying the problems of the single-layer perceptron as *rigorously* as they did (even though the existence of those problems was known long before they carried out their study) was indeed a contribution to neural networks.

On the multilayer network side, Minsky and Papert's criticism can also be seen as a positive contribution to the evolution of neural network research, although of a different kind. With their famous challenge about the problem of learning in multilayer neural networks, they directed the attention of future researchers to this problem. The problem was well known for early neural network researchers, but after Minsky and Papert's (1969) study its "reverse salient" (in Hughes' terms) character became even clearer, because of the importance of that study for the closure of the perceptron controversy. By being published so late in the perceptron controversy, Minsky and Papert's "Perceptrons" study had an interesting effect. Although in the short term the study was the "last word" in the controversy, in the long term some of the weak points of neural network research became clearer after it. In particular, the problem of training multilayer systems remained as a challenge for future neural network researchers.

In the early 1980s, the new generation of neural network researchers did indeed take Minsky and Papert's "pessimistic intuitive judgement" about that problem as a challenge, and developed important solutions for it (I will come to this in Section 11). But, because of the importance of Minsky and Papert's study for the closure of the perceptron controversy, it would not be surprising to find that someone had tried to solve that problem before Rumelhart and his colleagues developed their technique in the mid-1980s.

In fact, Rosenblatt himself was quite close to the idea of back-propagation.

> . . . 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)

It has been pointed out that techniques of some similarity with back-propagation were being used in control theory by the 1950s (le Cun, 1988;

Hecht-Nielsen, 1990).[39] Nevertheless, it seems that the first attempt to apply something similar to back-propagation to neural network research was Paul Werbos'. Werbos works now at the National Science Foundation (NSF), where he is responsible for the funding of some neural network projects. In the early 1970s, Paul Werbos was carrying out Ph.D. research in applied mathematics at Harvard University. He 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 minimizing a total error function is to calculate the contributions of the internal (or 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. In the 1970s, Werbos developed a technique that he called "dynamic feedback" as a solution to that problem (Werbos, 1974). The idea was to propagate information backward along the network, so that the derivatives of the error with respect to the intermediate units could be calculated. Werbos acknowledged that ideas of some similarity with his "dynamic feedback" had been used in control theory earlier.[40]

Werbos' attempts at applying his dynamic feedback method to neural network-like systems found strong resistance in the scientific community. At that time, AI was dominated by the symbol-processing approach, and neural network-like techniques were not popular. Werbos believes that some of the

[39] "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 back-propagation] recursive control algorithm was presented by Arthur Bryson and Yu-Chi Ho . . . in 1969. The primary learning law used can be shown to follow from the Robbins/Monro technique introduced in *1951 . . . The earliest incarnation of* backpropagation has probably not yet been found" (Hecht-Nielsen, 1990, pp. 124–125).

[40] "Werbos (1974) also cited related work in control theory, which also used backwards flows of information to identify systems, albeit in a different way. [My] 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 back-propagation] was the use* of dynamic feedback in combination with those two components" (Werbos, 1988, p. 341).

difficulties he had throughout his doctoral research at Harvard University were related to that. The members of his thesis committee had doubts about the validity of his work, and he was told to talk to someone with enough expertise and credibility. This is when (the early 1970s) 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 do derivatives, and lets you make these things work, and that overthrows what you did in your Perceptrons [Minsky and Papert, 1969] book.' Do you know how enthusiastic Minsky was about that? I believe it was 1970 or 1971 when I presented it to Minsky, and I remember his reaction very vividly. I think that 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 thing. I am not afraid to name Minsky because everyone will continue to value his contributions, regardless. His theorems—though misused in the past—are still of use today, and I cite them in my most important work in the Handbook of Intelligent Control [White and Sofge, eds., 1992].' (Werbos, interview)

Minsky told me about Werbos accidentally, without having been asked about him. I told him that some people 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. This was his reply:

> '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, in the early 1970s, Minsky thought that the idea of back-propagation was not practical (he still has many doubts about it today, as I will show later). But what interests me most in this section is that Werbos was unable to "sell" his idea in the AI community. The idea of back-propagation is important within a neural network context, but in the 1970s (after the closure of the controversy) neural network research simply did not count as a legitimate approach to AI.

But let me come back to Werbos' story. As time went by in the early 1970s, and pressure from the Harvard thesis committee was rising, he wrote a simpler and clearer paper about the back-propagation idea. Werbos claims that 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 Ph.D. 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)

Finally, K. 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 (see Werbos, 1988, p. 342). 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.[41]

[41] "'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 in the thesis 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 [of Werbos, 1974] which are still useful in the neural net profession today" (Werbos, interview).

Later, Werbos made further attempts at applying his technique to multilayer neural networks. In the early 1980s he was working at the U.S. Department of Energy under Charles Smith's direction, where he had carried out some nonneural applications of back-propagation. According to Werbos, he applied to Smith for a project aiming at applying his dynamic feedback technique to neural networks, but without success. Werbos complains that Smith was afterward among the people who funded the PDP group, the group that finally developed the back-propagation technique within the neural network context.[42]

Werbos thinks that the paper he presented at the 1981 International Federation for Information Processing (IFIP) conference in New York (Werbos, 1982)—a condensed version of an EIA paper he had written for Smith—guarantees his priority claims about back-propagation. In that paper, there were some comments about the possibility of applying this algorithm to neural network systems (see Werbos, 1982, p. 765). After the paper was published, Werbos felt that his priority was guaranteed, and tried hard to communicate and extend his idea. In sum, Werbos claims that the idea of back-propagation 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 the relevant places . . . In 1982 [after the IFIP paper was published] I was very free and open, and did my best to push the nondocumented communication. Thus I feel it is no coincidence that the method was suddenly rediscovered in three places in the mid-1980s, two of them being places I had directed my efforts towards. (Werbos, interview)

Priority controversies are a classic theme in the sociology of science.[43] These disputes are a consequence of the social organization of science. Scientific

[42] "I had carried out a successful application of back-propagation under Charles Smith's authority . . . [After that, one day] I gave him [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. 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).

[43] See, for example, Hagstrom (1965) and Merton (1973). W. Hagstrom pointed at 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 recognize 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" (Hagstrom, 1965, pp. 14–15).

knowledge is produced and validated in a system of communication and control in which recognition is given as a reward for novelties (discoveries, inventions, new developments). Recognition is a sort of symbolic "capital," which is used by scientists to carry out further research (e.g., larger or more important projects). But the notion of discovery is problematic. In its usual sense, this term refers to a single, discrete act, localizable in time and space. 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"). After carrying out several case studies, Robert Merton problematized this notion, and claimed that all scientific discoveries are "multiples."[44] But one could go even further and affirm that the "discovery" or "non-discovery" of a new result or technique is (at least sometimes) linked to the controversy about the *validity* of that result, experiment, or technique (see Collins, 1985, p. 89; Barnes, 1982, p. 45).

Werbos' claim about the influence of his work on the development of back-propagation by Rumelhart, Hinton, and Williams will have to remain as an allegation. In a social activity such as science, the interactions and relationships among the actors involved are of multiple kinds and go in many different directions. Furthermore, a good deal of the knowledge developed, used, transmitted, and transformed in the interactions among scientists is tacit, and this makes priority disputes bitter and difficult to solve.

What is interesting from the point of view of this chapter is that Werbos was unable to overcome the resistance that he found to the *very idea* of applying a technique similar to back-propagation in AI research. Werbos' technique could have been used as an argument in favor of neural networks in the 1970s, but AI was very much dominated by the symbol-processing approach, and neural networks were still a paradigm without scientific value. Within this context, it is not surprising that Werbos was unsuccessful in his attempts at applying his dynamic feedback technique to neural networks. The anti-neural network position remained unchallenged until the mid- and late 1980s, when neural networks re-emerged as an AI paradigm in its own right.

11. Back-Propagation: Learning in Multilayer Perceptrons

A good part of the work done in neural networks in the 1980s was directed toward solving the problem of training multilayer systems. Building on work

[44] ". . . 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).

by Hopfield (1982) and others, Ackley, Hinton, and Sejnowski, from the PDP group, developed a first solution to that problem, the so-called Boltzmann machine learning algorithm (Ackley *et al.*, 1985).[45] The architecture of their network (with symmetric connections) was different from the usual feedforward one. The Boltzmann learning algorithm was rather slow (it is currently being developed in different directions to make it faster), but it encouraged researchers to develop other techniques for multilayer networks. In 1986, Rumelhart, Hinton, and Williams, from the PDP group, developed the back-propagation technique. They derived their back-propagation learning equations as a generalization of Widrow and Hoff's (1960) weight-modification algorithm for single-layer networks using the chain rule for differentiation (Rumelhart *et al.*, 1986a, pp. 322–328).

Figure 12 shows the architecture of the neural network studied by Rumelhart, Hinton, and Williams (1986a). 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 feed-forward way. 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). Unlike in Rosenblatt's perceptron, in a multilayer network all the connections have modifiable values. The back-propagation network of Fig. 12 is very similar to some networks studied by Widrow and his colleagues in the 1960s (see Section 4). However, there are important differences too, and these differences were also important for the development of the back-propagation algorithm.

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. Rumelhart and his colleagues used a continuous, differentiable, sigmoid threshold function instead of the step function used by early neural network researchers. Thanks to this change, the reverse salient of training multilayer nets could be redefined as a problem that could be solved.

[45] Hinton and his colleagues developed John Hopfield's (1982) neural network model further to give a first solution to the problem of learning in multilayer systems. The crucial aspect of Hopfield's 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 minimization of a *global* 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.

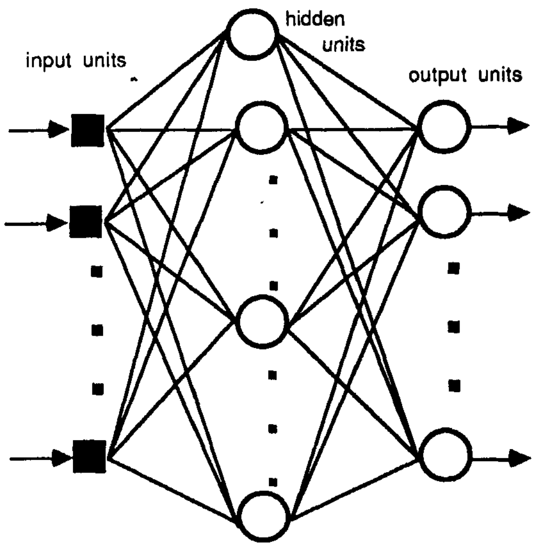

FIG. 12. Multilayer perceptron.

> It is really interesting to reread Nilsson's book on early neural networks [Nilsson, 1965] because you can see the exact assumptions, often they are very small assumptions, having to do with how you define the input–output function. 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)

What one needs to know in order to adjust the weights in a multilayer back-propagation 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 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 should depend on the errors made by the output units to which that unit is connected. These errors are back-propagated, so that the weights between input units and hidden units can then be adjusted. In a back-propagation network, each output unit demands from the hidden units exactly what it needs, and the hidden units try to accommodate the conflicting demands.

A learning cycle in a back-propagation network can be summarized as follows. A pattern p is presented, activity propagates forward throughout the units, and the network produces an output. This output is compared with the desired output, and the error made by the output units is calculated. Then, before any weight adjustment is made, the backward stage starts. The errors made by the output units are back-propagated to the hidden units, so that the error made by each hidden unit can be calculated. Now all the connections in the system can be changed. 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. By adjusting the connections of the system according to the back-propagation technique, the total error measure for a set of input/output patterns is minimized in a gradient descent way.[46]

Multilayer systems with back-propagation became the most popular neural networks of the late 1980s. Rumelhart, Hinton, and Williams' back-propagation technique was an important element for the reopening of the neural network controversy. Of course, this importance has to be understood within the context of emergence of neural networks in the 1980s. (I discussed some of the main aspects of this context in Section 9.)

At the same time as Rumelhart, Hinton, and Williams developed their back-propagation network in the mid-1980s two other researchers—Parker (1985) in the United States and le Cun (1985) in France—developed similar techniques. There have been no open priority disputes between le Cun, Parker, and Rumelhart and his colleagues, and it is usually accepted that the success of introducing back-propagation in neural computing was due to Rumelhart, Hinton, and Williams. According to Rumelhart, their work on back-propagation was carried out independently from Parker, le Cun's, and (importantly) Werbos'.

> 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)

[46] "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 *et al.*, 1986b, p. 697).

> 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. (Rumelhart, interview)

Within the context of the re-emergence of neural network research in the 1980s, the reaction to back-propagation within the emerging neural network community was rather optimistic. Back-propagation was seen by many neural network researchers as the most important technical result of the mid- and late 1980s. For the growing neural network research community of the late 1980s, back-propagation was a successful solution to the historical reverse salient of training multilayer perceptrons.

A great part of the importance of the back-propagation algorithm is due to the classification power of multilayer neural networks. This has been an attractive property of multilayer systems since the early 1960s (see Hawkins, 1961, p. 47). After Rumelhart and colleagues' results on back-propagation, the classification power of multilayer networks received increasing attention. Classifications realized by neural networks can be represented as decision regions in pattern space. Multilayer neural networks with two layers of hidden units and three layers of modifiable connections can form any decision region in pattern space, that is they can realize 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), i.e., any input/output function (or classification). (Remember that the back-propagation weight modification algorithm applies equally to a network with three layers of connections.)

But even though it is very important, "theoretical" classification power is only one issue in neural computing. Other very important issues are the number of learning cycles required for a classification task, and the number of units needed. It will be seen that Rumelhart and colleagues' claims about their back-propagation results and experiments were pretty much about questions of practice.

From the point of view of this chapter, it is important to note that, 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 (their "intuitive judgement") about multilayer networks. Rumelhart, Hinton, and Williams claimed that their back-propagation learning algorithm was a successful response to Minsky and Papert's challenge. They acknowledged that sometimes the network gets trapped in local (or false) minima of the error landscape, but 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 guaranteed 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 guaranteed 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 *et al.*, 1986a, pp. 321, 324, 361)

Thus, Rumelhart and his colleagues claimed that they had given a response to Minsky and Papert's challenge about the possibility of developing a learning algorithm for multilayer systems. But Minsky and Papert "counter-attacked" and criticized Rumelhart and colleagues' optimistic evaluation of back-propagation (Minsky and Papert, 1988). The difference between Minsky and Papert's evaluation of the back-propagation results and Rumelhart and colleagues' original claims is striking. This is 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 generalized delta rule, or back-propagation] 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 [back-propagation'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 and Papert, 1988, pp. 260–261)

I asked Minsky about the claim by Rumelhart and his colleagues that Minsky and Papert's "intuitive judgement" about training multilayer systems was misplaced and that an effective training technique for multilayer perceptrons has now been developed.

> The book [Minsky and 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 [back-propagation 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 fuzziness. (Minsky, interview)

I discussed at the end of Section 5 what Minsky and Papert *said* on the problem of training multilayer networks. However, what Minsky and Papert said was widely *interpreted* as showing that learning in multilayer systems was a very important problem and that the neural network approach was not worth pursuing. In the quotation just given, Minsky suggests that many of his 1969 conclusions still hold. He points out that, even though the concept of efficiency has changed (because of developments in computer technology), "we don't know if back-propagation networks solve any difficult problems." But the way Rumelhart and his colleagues see the problems of hill-climbing (or gradient descent) techniques is remarkably different.

> ... The procedure we have produced [back-propagation] 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 *et al.*, 1986a, p. 328)

These disagreements in the evaluation of the back-propagation technique between Minsky and Papert on the one hand and the PDP researchers on the other are another case of interpretative flexibility (i.e., of divergent views about 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 small-scale examples studied by Rumelhart and his colleagues in their 1986 paper.[47] Thus, 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.

In the late 1980s, problems like parity and connectedness were again a matter of controversy between Minsky and Papert and neural network researchers (PDP researchers this time round), as they had been at the time of earlier neural network research (see last footnote). There are certain parallels between Minsky and Papert's (1988) criticism of back-propagation and their earlier criticisms of the perceptron and neural networks. In 1988, Minsky and Papert claimed that problems like parity had not been successfully solved, and they also criticized the adequacy of back-propagation—and gradient descent methods in general—for AI research, as they had done earlier (remember the Werbos story).

[47] It is interesting to look at Minsky's discussion of the small examples used by Rumelhart and his colleagues in their 1986 paper. Rumelhart, Hinton, and Williams (1986a, pp. 334–335) described a small network that was able to compute parity. The (mini) network had four input units, four hidden units, and one output unit. Since it had 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 showed that back-propagation was a successful learning algorithm for multilayer networks. Minsky and Papert 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" (Minsky and Papert, 1988, p. 254). I asked Minsky about the parity problem. "We don't have any theory of what range of problems they [back-propagation 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).

Minsky and Papert's earlier criticism was very important in the crystallization of the consensus against neural networks in the late 1960s, but their renewed criticism did not have the same effect in the late 1980s. Their renewed arguments were not seen as decisive against neural networks in the context of re-emergence of neural network research in the late 1980s. The neural network controversy was reopening, and this process could not be stopped. Arguments that had a strong weight in the late 1960s were not so important in the new context. In spite of Minsky and Papert's (1988) renewed criticism, many researchers now see gradient descent methods like back-propagation as perfectly adequate in AI research, and Rumelhart and colleagues' back-propagation results are widely seen as an adequate solution to the historical "reverse salient" of training multilayer networks. Of course I do not mean that Minsky and Papert's (1988) renewed criticism was misplaced or exaggerated (my purpose here is not to evaluate it). This criticism, like the 1969 one, created (constructed) problems for neural network research, and therefore it was in a sense a positive contribution to the development of the connectionist paradigm.

Nowadays, back-propagating multilayer networks have become a subarea of research in its own right, with an increasing number of researchers dedicating important efforts to developing, improving and applying them.[48] One of the earliest important applications of back-propagation was Sejnowski and Rosenberg's NETtalk network, a multilayer neural network with back-propagation that learns to pronounce an English text (Sejnowski and Rosenberg, 1986, 1987). In one of their experiments, Sejnowski and Rosenberg used a training set consisting of 1,024 words. They reported that (1986, p. 665), 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. Afterward, the network's capacity for generalization was tested by presenting to it a 493-word continuation from the same speaker without training. Sejnowski and Rosenberg (1986, p. 667) reported an 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 with DECtalk (from Digital Equipment Corporation), at that time 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.

[48] For a review on some attempts to minimize the occurrences of local minima when using back-propagation see Beale and Jackson (1990, pp. 91–97).

DECtalk (based on D. Klatt's work) took about 15 years of research to develop. For neural network researchers, NETtalk was an example of the advantages of neural computing. However, D. Klatt recently "counter-attacked" by claiming that the (acoustic–phonetic rule-based) alphabetic-to-phonetic representation system of DECtalk took only three years to develop (DARPA, 1988, p. 219). NETtalk was not pursued further toward commercialization (p. 220), but it is a good example of the early, and for many researchers promising, applications of networks with back-propagation.

Some of the early applications of back-propagation were summarized in DARPA's 1988 neural network study. According to this study, back-propagation 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 back-propagation 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" (DARPA, 1988, pp. 203–221).

Since then, multilayer back-propagation perceptrons are being applied to a great variety of problems. A discussion of these applications is out of the scope of this chapter. For a review of some neurobiology-oriented applications, see Anderson (1988). Le Cun *et al.* (1989) applied multilayer back-propagation 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), electrocardiograph 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 check signature verification. Another application, developed by Lehky and Sejnowski in 1988, is a neural network that 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 back-propagation networks in several problems. These include NETtalk-like systems for prediction of secondary structure of proteins (Qian and Sejnowski 1988), which allegedly outperformed the best available alternative method, and for hyphenation. Other application areas are 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 noncommercial) at the 1989 London "computer olympiad" (Hertz *et al.*, 1991, p. 137).

Most of these applications are in their early stages (research and development is still being done), and they have not reached commercialization yet. However, they show clearly that a considerable amount of research and development activity is now being carried out in neural computing. Although it is not yet clear how far the institutionalization 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 that 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 back-propagation learning algorithm in the context of re-emergence of connectionism was of great importance for the reopening of the controversy. In section 13, I will look at the new neural network debate, but before that I would like to review briefly some aspects of the "neural network explosion" of the late 1980s.

12. The Neural Network Explosion

In this section I will characterize briefly some aspects of the growth and institutionalization of neural network research in the late 1980s. I use the term "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 have become a research specialty and a research and development (R&D) community in its own right.

The growth of the neural network community can be measured in terms of conferences, publications, funding, and migration of scientists and engineers to the neural network field. Michel Rappa and Koenraad Debackere, from the MIT Alfred P. Sloan School of Management, carried out quantitative research of this kind (Rappa and Debackere, 1989, 1990).

After building a relational database containing 2,740 abstracts of journal and conference proceedings papers on neural networks published between 1969 and 1988, Rappa and Debackere (1989, pp. 9–10) concluded that the biggest increase in the rate of growth of the neural network research community happened from 1986 to 1988. The rate of growth of the community was 60% during those years: the community expanded from 200 members to 1,200. In a later study, Rappa and Debackere (1990) carried out a statistical survey of 700 neural network researchers from 30 countries. The survey was done between February and May, 1990, and it confirmed that the first peak in the entry of researchers to the field took place in 1986 and 1987. Rappa and his colleague concluded that 75% of their respondents had entered the field between 1984 and 1990, whereas only 25% had entered before 1984.

One of the characteristics of the emergence and institutionalization of research specialties is the proliferation of scientific conferences. This seems to be a good "thermometer" of the growth of neural network research also. The Santa Barbara (California) Neural Networks for Computing meeting, organized by the American Institute of Physics (AIP) in 1985, was one of the first meetings of the new neural network period, and it showed the interest of the physicists in the field. The migration of physicists to neural networks was a significant feature of the re-emergence of the field in the 1980s (Hopfield's case is an example of this). The Santa Barbara meeting had 60 participants (Denker, 1986, preface). One year later, in April 1986, the American Institute of Physics organized its second Neural Networks for Computing conference in Snowbird, Utah (Denker, 1986). 160 people attended the meeting, and the organizers were short of space to accept more participants.

There is an interesting anecdote about that 1986 Snowbird conference that illustrates some features of the emergence of neural networks in the 1980s. Bernard Widrow was "rediscovered" there—but he had to raise his hand and introduce himself, because nobody had recognized him.

> In the 1970s and 1980s we continued working on adaptive filtering and adaptive signal processing until about 4 years ago, when I heard about a meeting at Snowbird [on neural networks]. I went to the Snowbird conference. I found people there that were so enthusiastic about 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. There were 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 leading member of the new neural network community, and he therefore is one of the few researchers who have played an active role in the two main peaks of activity in the history of neural networks. In 1988, Widrow was the director of the DARPA neural network study (DARPA, 1988; after this study, DARPA decided to start funding some neural network projects), and in the same year he was appointed president of the International Neural Network Society (INNS).

1987 was the year of the first massive neural network conference. In June the Institute for Electrical and Electronic Engineers (IEEE) organized the First International Conference on Neural Networks (ICNN) in San Diego, California. This first ICNN was attended by 1,500 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 1,200 members. In the July 1987 annual meeting of the American Association of Artificial Intelligence (AAAI) a workshop on neural networks was organized for the first time, and some technology products related to neural networks were exhibited there. Since then a lot of conferences on neural networks have been organized. Rappa and Debackere (1989, p. 28) pointed out that in 1988 and early 1989 there were about 31 conferences on neural networks. One of the most important ones in 1988 was IEEE's second ICNN conference, attended by 2,200 people (*EE Times*, 1988, p. 49). A talk by Marvin Minsky seems to have been one of the most exciting moments of the conference.

> 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. (Zeitvogel, 1988a, pp. 10–11).

An important event that occurred at the 1988 ICNN was DARPA's announcement of its neural network program.

Another important conference held that year was the INNS's first annual conference. About 600 papers were presented there, twice the number of papers presented at the 1988 ICNN (Zeitvogel, 1988b, p. 12). Also in 1988, INNS created the Neural Networks journal, one of the most important publications in neural network research. By 1989, INNS had 3,500 members from 38 countries (Rappa and 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 legitimize neural network research. DARPA's support for neural network research was especially significant because of the agency's role in the development of symbolic AI throughout the years.[49] One of the three goals of the DARPA program 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. The other two goals were the development of neural network theory and modeling on the one hand, and neural network hardware implementation technology on the other (see DARPA, 1988, pp. 10–11).

By 1989, most of the major U.S. funding agencies had launched programs in neural networks. Those agencies include NSF, ONR, AFOSR, NASA, and NIH. Neural network programs were also launched by the European Community in 1988 and 1989, and by several European governments (the program 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 program will be a part of the Sixth Generation Computer Program to be launched after the Fifth Generation Project comes to an end in 1991.[50] Many major information technology companies in the U.S., Japan, and Europe, as well as small companies specialized in neural networks (mainly in the U.S.), are now developing neural network products. In December 1990 it was estimated that there were some 300 vendors of neural network technology products world-wide (Molina, 1990, p. 368), but the commercialization of neural computing is still in its early stages.

In the academic world, there are an increasing number of Ph.D. 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 centers dedicated to neural computing are emerging. The disciplinary origins of current neural network researchers is another interesting aspect of the recent growth and institutionalization of the neural network community. Rappa and Debackere

[49] This legitimizing role of DARPA is recognized within the DARPA study itself, and compared with the agency's lack of support for neural networks in the past (DARPA, 1988, p. 23). 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, 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).

[50] For short reviews on government funding for neural networks in Japan, the United States, and Europe see Molina (1990) and Johnson and Schwartz (1990).

concluded that, in 1990, 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 and Debackere, 1990, p. 12).

Another feature of the institutionalization of neural network research is the appearance of scientific journals exclusively dedicated to it. Apart from the journal of the INNS, called *Neural Networks*, which debuted in 1988, an increasing number of specialized journals dedicated to neural computing have appeared more recently. These include two in 1989: *Neural Computation* (U.S., edited by T. Sejnowski; it 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 U.S.

Apart from this, most AI, cognitive science, electronic engineering, and many philosophy journals have published special issues on neural computing.[51] A number of books on the foundations of neural networks and collections of historical papers have also been published.[52] Books on neural networks for the general public have also come out.[53] Recent books on the foundations of AI, such as Graubard (1988) and Patridge and Wilks (1990), dedicate considerable attention to neural network research.

All these indicators (conferences, applications, publications, migration of researchers, and funding) show that the emergence of neural networks as a research community of its own is well under way. The first important peak in this process happened at the end of the 1980s, when an increasing number of researchers from various disciplines started to "migrate" toward the neural network area.

[51] Examples of these 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), and *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 also. These include Beale and Jackson (1990), Aleksander and Morton (1990), Hecht-Nielsen (1990), and Hertz *et al.* (1991).

[52] Such as Nadel *et al.* (1989) and Anderson and Rosenfeld (1988).

[53] Examples of these are Johnson and Brown (1988), Allman (1989), and Brunak and Lautrup (1990).

13. The Current Debate: Conclusions

13.1 Debate Continues

In the previous section I showed that the process of emergence and institutionalization of neural network research is well under way. Neural network research is now generally accepted as an approach to AI and cognitive science in its own right. However, there is still quite a lot of debate about the exact "place" of neural network research in AI. This debate is going on right now, and it is very much open. After being closed in the 1960s, the neural network controversy has now been reopened. In this section I will review some of the main positions of the new controversy then make some concluding comments.

The three main positions (three interpretations of neural networks) that were emerging at the end of the 1980s I will call implementationism, moderate connectionism, and radical connectionism. The implementationist position is the most negative view of neural network research. Researchers favoring this position maintain 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 most a theory about how symbol-processing could be implemented in some kind of nonsymbolic substratum. Broadbent (1985) was one of the first researchers who espoused this interpretation of neural networks in his criticism of a paper by McClelland and Rumelhart (1985), in which these researchers advocated a distributed model of memory. More recently, cognitive science researchers Jerry Fodor and Zenon Pylyshyn (1988) made a strong defense of the implementationist hypothesis in a paper that created a good deal of controversy. Fodor and Pylyshyn situated their discussion at the level of representation and algorithm (on the notion of "information processing levels" [see Marr, 1982, ch. 1]). They claimed that connectionism is irrelevant at that level, 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 and 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 logicosyntactic structure of cognitive processes. Their view is that the neural network approach cannot account for some of the basic elements of human cognition like compositionality. Thoughts and mental states have a compositional structure, and cognitive processes depend on that structure.[54] Fodor and Pylyshyn 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 that connectionist models carry out is based on statistical association, which is rather different from the formally or syntactically driven inferential processes favored by the symbolic approach.[55] Fodor and Pylyshyn (1988, p. 68) doubted whether association is useful *at all* in studying and modeling cognitive processes.

For Fodor and Pylyshyn, neural networks would be at the most 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 spreading activation among semantically interpreted nodes. (Fodor and Pylyshyn, 1988, pp. 67–68)[56]

[54] 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 different symbolic expressions (or composite symbol structures). Nevertheless, the principle of compositionality has been criticized 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).

[55] But 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.

[56] But see previous footnote.

It does not seem likely that neural network researchers are going to follow Fodor and Pylyshyn's appeal and "rewrite much of the polemical literature" or "redescribe their systems as operating on symbol structures." Neural network researchers have already won recognition for their paradigm as an approach to AI of its own, and accepting Fodor and Pylyshyn's suggestions would mean going backward. It is interesting to note that the implementationist interpretation of neural networks is often linked—as it is in Fodor and Pylyshyn's case—with a view according to which "nothing has changed in AI and cognitive science in spite of the recent developments in neural network research." This type of hypothesis is the most critical interpretation of neural network research.

> [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 *punctuate 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 and Pylyshyn, 1988, pp. 6, 49)

Fodor and Pylyshyn are not the only ones who have espoused this "nothing has changed" interpretation of neural network research. Leading researchers like Minsky and Papert (1988, p. vi), Tomaso Poggio (a leading machine vision researcher from MIT's AI laboratory and Thinking Machines Corporation), and Daniel Hillis (the computer scientist who developed the Connection Machine parallel computer) have made some remarks in that direction.[57] The openness of the current debate is quite clear when one looks at the "implementationist" or "nothing has changed" positions.

[57] "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). ". . . 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

Neural network researchers have replied to the implementationist criticisms in several ways. Here I will only look at a few of them. 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 argued 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 (see Smolensky, 1987, p. 151).

Other responses to Fodor and Pylyshyn criticized their assumption about the "independence of information-processing levels" (and particularly the independence between the symbol-processing level and the hardware level).[58] The degree of neurobiological constraint in neural network architectures varies considerably from system to system, but the interaction between neural network research and neuroscience is probably going to be one of the main neural network research areas in the future.[59]

The second position in the current debate about neural network research that I would like to look at briefly in this section could be called radical connectionism. Researchers in favor of this view claimed that neural networks will offer an alternative and sufficient way of explaining and modeling most cognitive processes. For them, the symbolic paradigm would only be a useful approximation to the connectionist description—the "right" description—of cognitive processes.[60] PDP researchers did not deny the

[58] ". . . 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, 1990, p. 94).

[59] This interdisciplinary research area is sometimes called "computational neuroscience" or "cognitive neuroscience." See Churchland and Sejnowski (1988) and Sejnowski *et al.* (1988).

[60] The following comments are examples of this radical connectionist position. "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

off, so the possibility that it might spontaneously emerge from the interactions of a large collection 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" (Hillis, 1989, pp. 175–176).

importance of "symbolic" concepts such as consciousness, sequential thought, and mental models, but they claimed that these phenomena could be modeled with purely connectionist systems (Rumelhart *et al.* 1986c). (For recent developments in neural network studies of symbol-processing, see *Artificial Intelligence*, 1990.)

Of course it remains to be seen to what extent the neural network approach is going to offer useful tools for the study and modeling of the more sequential, structure-sensitive cognitive and intelligent processes. Radical connectionism poses a research agenda for neural network research for the years to come. Furthermore, radical connectionist pronouncements have a rhetorical element of promise. However, one should not forget that the rhetoric of promise does play a role in science. Kuhn's comments about the function of the rhetoric of promise in the early stages of the evolution of a line of research seem to apply here.

> ... 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 ... (Kuhn, 1970, pp. 157–58)

The third and last position I would like to examine here is what I call the moderate connectionist view, a more eclectic view of the current debate between connectionism and symbolic AI. One of the researchers who has elaborated this position most explicitly is Andy Clark (1989a, 1989b), a philosopher from the School of Cognitive and Computing Sciences of the University of Sussex (Brighton, England). Clark defended hybrid (partly symbolic, partly connectionist) systems. He 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 adequate models, and not only "approximations" (contrary to what radical

level? The position taken by the subsymbolic [i.e., neural network] paradigm is: No—it lies at the subconceptual level" (Smolensky, 1988, p. 7). "... 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 *et al.*, 1986c, p. 56).

connectionists would claim).[61] Clark maintained that, even though from an evolutionary point of view neural network-like systems existed earlier (connectionist-like systems—the brain—had to learn to simulate a von Neumann-style architecture for tasks like addition or subtraction), what matters is not the evolutionary origin of cognitive faculties and processes, but the way those processes "function" (Clark, 1989a, p. 63, 1987, p. 13).

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 (1988) "neural phonetic typewriter" speech recognition system. The central part of the system is an unsupervised neural network that classifies phonemes. However, the preprocessing part is based on conventional digital signal processing techniques, and the postprocessing part is a symbolic rule base.[62]

So far in this section I have discussed briefly three positions of the current debate about neural networks. In spite of "nothing has changed" views, neural network research has been accepted as an approach to AI and cognitive science in its own right, but the controversy provoked by its re-emergence has not been closed. The definition of the relationships between the symbol-processing and the neural network approaches to AI and cognitive science is right now a matter open to controversy and negotiation. The current debate about the place of neural network research within AI and cognitive science (the current, in a sense, "territorial dispute") will shape these disciplines for years to come. The future shape of the map of AI and cognitive science is difficult to predict, but one thing seems clear: this time around the new neural network controversy is not going to end with the total rejection of one of the contending positions.

13.2. Conclusions

In this chapter I analyzed the main developments of the history of neural network research from a sociological and historical point of view. I have

[61] ". . . 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).

[62] Kohonen (1990, p. 1477) warned against insisting too much on distribution, and forgetting localization and organization of information into separate parts.

shown that the controversy-closure scheme from the sociology of science (Collins, 1981, 1985) can help interpret those developments. The neural network controversy was once closed (in the late 1960s) and has recently reopened.

The controversy of the 1960s ended with the rejection of the neural network approach (see Sections 2 through 8). Neural networks had many technical problems (including limitations of single-layer machines, and the lack of algorithms for training multilayer systems), but the importance of those problems was a matter of controversy (see Section 5). There were important technological limitations too, like the lack of computing power to simulate large networks efficiently. However, there was not a logically necessary connection between these problems and the rejection of neural networks *as a whole.* This rejection, i.e., the conclusion that *the whole* neural network paradigm lacked scientific validity, was a consequence of the closure of the perceptron controversy. That process of closure was a social, contingent process, and therefore its result (i.e., the total rejection of neural networks) was not the only possible result (at least in principle). Factors like the emergence and institutionalization of symbolic AI help explain the radical rejection and abandonment of neural networks (see Section 8). In a different context, neural networks re-emerged as a legitimate approach to AI in the late 1980s (see Sections 9 through 13). The consensus that emerged at the end of the perceptron controversy in the 1960s was revised, and controversy reopened.

But having said this, I want to make clear that by analyzing the neural network controversy sociologically I am not condemning any of the positions of the participants involved. As I said in Section 1, science is often generated and validated through controversies, and therefore debate can be seen as a positive force in the development of scientific knowledge. Controversy can be bitter at times, and the rhetoric used can be quite strong, but that is how science is produced, validated, and developed.

Appendix 1. 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 at San Diego, La Jolla, California, November 8, 1989. **Cicourel**, A., Cognitive Science, University of California at 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 at 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, Cambridge, Massachusetts, October 25, 1989. **Nilsson**, Nils J., Stanford University, Stanford, California, November 3, 1989 (by telephone). **Norman**, Donald A., University of California at San Diego, La Jolla, California, November 8, 1989. **Papert**, Seymour A., Massachusetts Institute of Technology, Cambridge, Massachusetts, December 4, 1989. **Reece**, 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, D.C., 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, D.C., 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 at San Diego, La Jolla, California, November 9, 1989.

Appendix 2. List of Personal Communications by Letter

Gwin, Cecil W., Martins Ferry, Ohio. **O'Brien**, Richard D., University of Massachusetts at Amherst. **Rosenblatt**, Maurice, Washington D.C. **Scattergood**, Mark, Englewood, Colorado.

Abbreviations Used

AFOSR: Air Force Office of Scientific Research
AI: artificial intelligence
DARPA: Defense Advanced Research Projects Agency (U.S.A.)
IEEE: Institute for Electrical and Electronic Engineers
ICNN: International Conference on Neural Networks
INNS: International Neural Network Society
LMS: least mean square

MIT: Massachusetts Institute of Technology
NASA: National Aeronautics and Space Administration
NIH: National Institutes of Health
NSF: National Science Foundation
ONR: Office of Naval Research (U.S.A.).
PDP: parallel distributed processing
SRI: Stanford Research Institute

Acknowledgments

The work at the basis of this chapter was supported by a Basque Government scholarship for doctoral research at the University of Edinburgh (Scotland) from 1988 to 1991. I would like to thank the following people for advising, encouraging, and helping me throughout my research: Donald MacKenzie (University of Edinburgh), Jesús María Larrazábal and Jesús Ezquerro (Universidad del País Vasco), Miguel A. Quintanilla (Universidad de Salamanca), and Josetxo Beriáin (Universidad Pública de Navarra). I am grateful to my American friends Joel Kallich and Sue Jennings, too, for helping me in many ways during my stay in the United States. I would also like to thank both the people I interviewed during my research (see Appendix 1) and those who provided me with valuable information by letter (see Appendix 2) for their time, attention, and interest in my work. Many thanks to Rafael Pardo (Universidad Pública de Navarra), too, for encouraging and helping me in my current research on the social dimensions of information technology. Finally, I would like to thank Txarete Ganboa for helping and supporting me in my research.

References

Ackley, D. H., Hinton, G. E., and Sejnowski, T. J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science* **9,** 147–169. (Reprinted in J. A. Anderson and E. Rosenfeld eds., 1988, pp. 638–649.)

Aleksander, I., and Morton, H. (1990). "An Introduction to Neural Computing." Chapman and Hall: London.

Allman, W. F. (1989). "Apprentices of Wonder: Inside the Neural Network Revolution." Bantam Books: New York.

Alternative Computers (1989). "Alternative Computers." Time-Life Books: Alexandria, Virginia.

Anderson, A. (1988). Learning from a computer cat. *Nature* **331,** 657–658.

Anderson, J. A., and Rosenfeld, E. (1988). "Neurocomputing: Foundations of Research." MIT Press: Cambridge, Massachusetts.

Arbib, M. A. (1989). Schemas and neural networks for sixth generation computing. *Journal of Parallel and Distributed Computing* **6,** 185–216.

Artificial Intelligence (1990). Special issue on connectionist symbol processing. *Artificial Intelligence* **46,** 1–2.

Ballard, D. H., Hinton, G. E., and Sejnowski, T. J. (1983). Parallel visual computation. *Nature* **306,** 21–26.

Barnes, B. (1982). "T. S. Kuhn and Social Science." The Macmillan Press Ltd.: London.

Barrow, H. G. (1989). AI, neural networks, and early vision. *AISB Quart.* **69,** 6–25.

Beale, R., and Jackson, T. (1990). "Neural Computing: An Introduction." Adam Hilger: Bristol, England.

Bernstein, J. (1981). Profiles: AI, Marvin Minsky. *The New Yorker* **December 14**, 50–126.
Block, H. D. (1962). The perceptron: a model for brain functioning, I. *Rev. Modern Phys.* **34**, 123–135. (Reprinted In J. A. Anderson and E. Rosenfeld, eds.), 1988, Neurocomputing: Foundations of Research, pp. 138–150. The MIT Press: Cambridge, Massachusetts.
Block, H. D. (1970). A review of 'Perceptrons'. *Information and Control* **17**, 510–522.
Block, H. D., Knight, B. W., and Rosenblatt, F. (1962). Analysis of a four layer series-coupled perceptron. *Review of Modern Physics* **34**, 135–142.
Broadbent, D. (1985). A question of levels: comment on McClelland and Rumelhart. *Journal of Experimental Psychology: General* **114** (2), 189–192.
Brunak, S., and Lautrup, B. (1990). "Neural Networks: Computers with Intuition." World Scientific: Singapore.
Chater, N., and Oaksford, M. (1990). Autonomy, implementation and cognitive architecture: a reply to Fodor and Pylyshyn. *Cognition* **34**, 93–107.
Churchland, P. S., and Sejnowski, T. J. (1988). Perspectives on cognitive neuroscience. *Science* **242**, 741–745.
Clark, A. (1987). Connectionism and cognitive science. In "Advances in Artificial Intelligence: Proceedings of the 1987 AISB Conference, University of Edinburgh, 6–10 April." (J. Hallam and C. Mellish, eds.), pp. 3–15. John Wiley and Sons: Chichester, Great Britain.
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". The MIT Press: Cambridge, Massachusetts.
Collins, H. M. (1981). Stages in the Empirical Programme of Relativism. *Social Studies of Science* **11**, 3–11.
Collins, H. M. (1985). "Changing Order: Replication and Induction in Scientific Practice." SAGE Publications Ltd.: London.
Congressional Record (1971). "Tribute to Dr. Frank Rosenblatt." United States Congressional Record: Proceedings and Debates of the 92nd Congress, July 28, 1971.
Cruz, C. A. (1988). "Understanding Neural Networks: A Primer." Graeme Publishing Corporation: Amherst, New Hampshire.
DARPA (1988). "Darpa Neural Network Study." Armed Forces Communications and Electronics Association (AFCEA) International Press: Fairfax, Virginia.
Darrach, B. (1970). Meet Shakey, the first electronic person. *Life* **69** (21), November 20, 58b–68.
Denker, J. S. (1986). "Neural Networks for Computing." American Institute of Physics: New York.
Dickson, D. (1988). "The New Politics of Science." The University of Chicago Press: Chicago.
Dreyfus, H. L., and Dreyfus, S. E. (1988). Making a mind versus modeling the brain. In "The Artificial Intelligence Debate: False Starts, Real Foundations." (S. R. Graubard, ed.), pp. 15–43. The MIT Press: Cambridge, Massachusetts.
EE Times (1988). Smart thinking shows at neural conference. *Electronic Engineering Times* **August 15**, 49, 65, 68.
Feldman, J. A., and Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science* **6**, 205–254. (Reprinted In J. A. Anderson and E. Rosenfeld, eds., 1988, pp. 484–507.)
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 M.Sc. dissertation, University of Manchester.
Fleck, J. (1982). Development and establishment in artificial intelligence. In "Scientific Establishments and Hierarchies, Sociology of the Sciences, Vol. VI." (N. Elias, H. Martins, and

R. Whitley, eds.), pp. 169–217. D. Reidel Publishing Company: Dordrecht, The Netherlands.

Fleck, J. (1984). Artificial intelligence and industrial robots: an automatic end for utopian thought. In "Nineteen Eighty-Four: Science between Utopia and Dystopia: Sociology of the Sciences Vol. viii." (E. Mendelsohn and H. Nowotny, eds.), pp. 189–231. Reidel Publishing Company: Dordrecht, The Netherlands.

Fleck, J. (1987). Postscript: the commercialisation of artificial intelligence. In "The Question of AI." (B. P. Bloomfield, ed.), pp. 149–164. Croom-Helm: London.

Fodor, J. A., and Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: a critical analysis. *Cognition* **28,** 3–71.

Graubard, S. R. (1988). "The Artificial Intelligence Debate: False Starts, Real Foundations." The MIT Press: Cambridge, Massachusetts.

Hagstrom, W. O. (1965). "The Scientific Community." Southern Illinois University Press: Carbondale, Illinois.

Harvey, B. (1981). Plausibility and evaluation of 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.

Hecht-Nielsen, R. (1990). "Neurocomputing." Addison-Wesley: Reading, Massachusetts.

Hertz, J., Krogh, A., and Palmer, R. G. (1991). "Introduction to the Theory of Neural Computation." Addison-Wesley Publishing Company: Redwood City, California.

Hillis, W. D. (1989). Intelligence as emergent behavior; or the songs of Eden. In "The Artificial Intelligence Debate: False Starts, Real Foundations." (S. R. Graubard, ed.). pp. 175–189. The MIT Press: Cambridge, Massachusetts.

Hockney, R. W., and Jesshope, C. R. (1988). "Parallel Computers 2: Architecture, Programming, and Algorithms." Adam Hilger: Bristol, England.

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 J. A. Anderson and E. Rosenfeld, eds., 1988, pp. 460–464.)

Hughes, T. P. (1983). "Networks of Power: Electrification in Western Society. 1880–1930." Johns Hopkins University Press: Baltimore, Maryland.

Johnson, R. C., and Brown, C. (1988). "Cognizers: Neural Networks and Machines That Think." John Wiley and Sons, Inc.: New York.

Johnson, R. C., and Schwartz, T. J. (1990). IJCNN government panel: governments fund neural nets worldwide. *Neural Technology Update (formerly Synapse Connection)* **4** (2), 1, 5–7.

Kohonen, T. (1988). 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.

Kuhn, T. S. (1970). "The Structure of Scientific Revolutions." (second, enlarged ed.). The University of Chicago Press: Chicago, Illinois.

Latour, B. (1987). "Science in Action: How to Follow Scientists and Engineers throughout Society." Open University Press: Milton Keynes, England.

Laudan, L. (1977). "Progress and Its Problems: Toward a Theory of Scientific Growth." University of California Press: Berkeley, California.

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 "Proceedings of the 1988 Connectionist Models Summer School." (D. S. Touretzky, G. E. Hinton, and T. J. Sejnowski, eds.), Morgan Kaufmann, Inc.: San Mateo, California.

le Cun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation* **1,** 541–551.

Lehky, S., and Sejnowski, T. J. (1988). Neural network model for the cortical representation of curvature from images of shaded surfaces. *In* "Sensory Processing in the Mammalian Brain Neural Substrate and Experimental Strategies," J. Lund (Ed.), Oxford University Press: Oxford, England.

Lighthill, J. (1973). "Artificial Intelligence." Science Research Council (Great Britain): London.

Lippmann, R. P. (1987). An introduction to computing with neural nets. *IEEE ASSP Magazine* **4,** April, 4–22.

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." The MIT Press: Cambridge, Massachusetts.

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." W. H. Freeman and Company: New York.

McClelland, J. L., and Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General* **114** (2), 159–188.

McCorduck, P. (1979). "Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence." W. H. Freeman and Company: New York.

Merton, R. K. (1961). Singletons and multiples in science. *Proceedings of the American Philosophical Society* **105** (5), 470–486. (Reprinted In "The Sociology of Science: Theoretical and Empirical Investigations." R. K. Merton, ed.), pp. 343–370. The University of Chicago Press: Chicago, Illinois.

Merton, R. K. (1973). "The Sociology of Science: Theoretical and Empirical Investigations." The University of Chicago Press: Chicago, Illinois.

Minsky, M. L., and Papert, S. A. (1969). "Perceptrons: An Introduction to Computational Geometry." The MIT Press: Cambridge, Massachusetts.

Minsky, M. L., and Papert, S. A. (1988). "Perceptrons: An Introduction to Computational Geometry." (expanded ed.). The MIT Press: Cambridge, Massachusetts.

Molina, A. H. (1987). "The socio-technical basis of the microelectronics revolution: a global perspective." Ph.D. thesis. University of Edinburgh, Scotland.

Molina, A. H. (1990). Emerging neural computing in the U.S.A., Japan, and UK/Europe. *Science and Public Policy* **17** (6), 363–371.

Nadel, L., Cooper, L. A., Culicover, P., and Harnish, R. M. (1989). "Neural Connections, Mental Computation." The MIT Press: Cambridge, Massachusetts.

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. (1983). Intellectual issues in the history of artificial intelligence. In "The Study of Information: Interdisciplinary Messages." (F. Machlup, U. Mansfield, eds.), pp. 187–227. John Wiley & Sons: New York.

Newell, A., and Simon, H. A. (1976). Computer science as empirical enquiry: symbols and search. *Communications of the Association for Computing Machinery* **19,** 113–126. (Reprinted In "Mind Design." J. Haugeland, ed.,) The MIT Press: Cambridge, Massachusetts.

Newsweek (1958). Human brains replaced? *Newsweek*, July 21, 50.

Nilsson, N. J. (1965). "The Mathematical Foundations of Learning Machines." Morgan Kauffmann Publishers: San Mateo, California.
Nilsson, N. J., and Raphael, B. (1967). Preliminary design of an intelligent robot. In "Computer and Information Sciences-2." (J. T. Tou, ed.), pp. 235–259. Academic Press: New York.
Norberg, A. L. (circa 1990). "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.).
Olazaran, M. (1991). "A historical sociology of neural network research." Ph.D. dissertation, University of Edinburgh, Edinburgh, Scotland.
Papert, S. A. (1988). One AI or many? In "The Artificial Intelligence Debate: False Starts. Real Foundations." (S. R. Graubard, ed.), pp. 1–14. The MIT Press. Cambridge, Massachusetts.
Parker, D. B. (1985). "Learning-Logic (TR-47)." Center for Computational Research in Economics and Management Science, MIT, Cambridge, Massachusetts.
Partridge, D., and Wilks, Y. (1990). "The Foundations of Artificial Intelligence." Cambridge University Press: Cambridge, England.
Qian, N., and Sejnowski, T. J. (1988). Predicting the secondary structure of globular proteins using neural network models. *Journal of Molecular Biology* **202**, 865–884.
Raphael, B. (1976). "The Thinking Computer: Mind Inside Matter." Freeman: San Francisco.
Rappa, M. A., and 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: Cambridge, Massachusetts.
Rappa, M. A., and Debackere, K. (1990). "International Survey on the Neural Network Research Community." Massachusetts Institute of Technology, Alfred P. Sloan School of Management: Cambridge, Massachusetts.
Roberts, L. G. (1963). "Machine Perception of Three Dimensional Solids." (Technical report no. 315). MIT Lincoln Laboratory: Lexington, Massachusetts.
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 and E. Rosenfeld, eds. pp. 92–114, 1988.)
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. In "Mechanisation of Thought Processes." pp. 421–456. London: Her Majesty's 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." Spartan: New York.
Rosenblatt, F. (1962b). Strategic approaches to the study of brain models. In "Illinois Symposium on Principles of Self-Organization (University of Illinois, Urbana, Illinois)." (H. von Foerster and G. W. Zopf, eds.), pp. 385–396. Pergamon Press: New York.
Rosenblatt, F. (1964). A model for experimental storage in neural networks. In "Computer and Information Sciences." (J. T. Tou and R. H. Wilcox, ed.), pp. 16–66. Spartan Books: Washington, D.C..
Rosenblatt, F. (1967). Recent work on theoretical models of biological memory. In "Computer and Information Sciences-II." (J. T. Tou, ed.), Academic Press: New York. (Proceedings of the Second Symposium on Computer and Information Sciences held at Batelle Memorial Institute.)
Rosenblatt, F., Farrow, J. T., and Herblin, W. F. (1966). Transfer of conditioned responses from trained rats to untrained rats by means of a brain extract. *Nature* **209**, 46–48.
Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986a). Learning internal representations by error propagation. In "Parallel Distributed Processing: Explorations in the Microstructure

of Cognition. Vol. 1. Foundations." (D. E. Rumelhart, J. L. McClelland, and The PDP Research Group, eds.), pp. 318–362. The MIT Press: Cambridge, Massachusetts.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986b). Learning representations by back-propagating errors. *Nature* **323,** 533–536. (Reprinted In J. A. Anderson and E. Rosenfeld, eds., 1988, pp. 696–699.)

Rumelhart, D. E., Smolensky, P., McClelland, J. L., and Hinton, G. E. (1986c). Schemata and sequential thought processes in PDP models. In "Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 2. Psychological and Biological Models." (J. L. McClelland, D. E. Rumelhart, and The PDP Research Group, eds.), pp. 7–57. The MIT Press: Cambridge, Massachusetts.

Sejnowski, T. J. (1987). Computing with connections. *Journal of Mathematical Psychology* **31,** 203–210.

Sejnowski, T. J., and Rosenberg, C. R. (1986). "NETtalk: A Parallel Network That Learns to Read Aloud (JHU/EECS-86/01)." The Johns Hopkins University, Electrical Engineering and Computer Science Dept. (Reprinted In J. A. Anderson and E. Rosenfeld, eds., 1988, pp. 663–672.)

Sejnowski, T. J., and Rosenberg, C. R. (1987). Parallel networks that learn to pronounce English text. *Complex Systems* **1,** 145–168.

Sejnowski, T. J., Koch, C., and Churchland, P. S. (1988). Computational neuroscience. *Science* **241,** 1299–1306.

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.

Star, S. L. (1989). "Regions of the Mind: Brain Research and the Quest for Scientific Certainty." Stanford University Press: Stanford, California.

Tesauro, G. (1990). Neurogammon wins computer olympiad. *Neural Computation* **1,** 321–323.

The Economist (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. (1986). Frank Rosenblatt: Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. In "Brain Theory." (G. Palm and A. Aersten, eds.), pp. 245–248. Springer-Verlag: Berlin.

von Foerster, H., and Zopf, G. W. (1962). "Illinois Symposium on Principles of Self-Organization." (University of Illinois, Urbana, Illinois). Pergamon Press: New York.

Werbos, P. J. (1974). "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences." Ph.D. thesis, Harvard University, Cambridge, Massachusetts.

Werbos, P. J. (1982). Applications of advances in nonlinear sensitivity analysis. In "Systems Modelling and Optimization: Proceedings of the 10th IFIP Conference. New York City, USA, August 31–September 4, 1981." (R. F. Drenick and F. Kozin, eds.), Springer-Verlag: New York.

Werbos, P. J. (1988). Generalization of backpropagation with application to a recurrent gas market model. *Neural Networks* **1,** 339–356.

White, D., and Sofge, D. (1992). "Handbook of Intelligent Control: Natural, Adaptive, and Fuzzy Approaches." Van Nostrand.

Widrow, B. (1962). Generalization and information storage in networks of adaline "neurons". In "Self-Organizing Systems-1962." (M. C. Yovits, G. T. Jacobi, and G. D. Goldstein, eds.), pp. 435–461. Spartan Books: Washington, D.C.

Widrow, B., and Hoff, M. E. (1960). Adaptive switching circuits. In "1960 IRE WESCON Convention Record." pp. 96–104. IRE: New York.

Widrow, B., and 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., and Goode, B. (1967). Adaptive antenna systems. *Proceedings of the IEEE* **55**, 2143–2159.

Wilks, Y. (1990). Some comments on Smolensky and Fodor. In "The Foundations of Artificial Intelligence: A Sourcebook." (D. Partridge and Y. Wilks, eds.), pp. 327–336. Cambridge University Press: Cambridge, England.

Yovits, M. C., and Cameron, S. (1960). "Self-Organizing Systems: Proceedings of an Interdisciplinary Conference (Chicago 5–6 May 1959)." Pergamon Press: New York.

Yovits, M. C., Jacobi, G. T., and Goldstein, G. D. (1962). "Self-Organizing Systems 1962." Spartan: Washington, D.C.

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, 12.