• *ABSTRACT*

In this paper, I analyze the controversy within Artificial Intelligence (AI) which surrounded the 'perceptron' project (and neural nets in general) in the late 1950s and early 1960s. I devote particular attention to the proofs and arguments of Minsky and Papert, which were interpreted as showing that further progress in neural nets was not possible, and that this approach to AI had to be abandoned. I maintain that this official interpretation of the debate was a result of the emergence, institutionalization and (importantly) legitimation of the symbolic AI approach (with its resource allocation system and authority structure). At the 'research-area' level, there was considerable interpretative flexibility. This interpretative flexibility was further demonstrated by the revival of neural nets in the late 1980s, and subsequent rewriting of the official history of the debate.

A Sociological Study of the Official History of the Perceptrons Controversy

Mikel Olazaran

The recent sociology of scientific knowledge has shown that processes of controversy often play a central role in the production and validation of scientific knowledge.[1] Harry Collins recommended the study of 'interpretative flexibility' (that is, variation in scientists' perceptions of the same results or experiments) as a methodological starting point in controversy studies.[2] Before consensus is reached, groups of scientists from diverse traditions (with their own cultures, interests and connections with both the wider scientific community and the wider society) may interpret the same experiment (phenomenon, result, method or technique) differently. Showing the interpretative flexibility of scientific results amounts to the realization that no knowledge possesses absolute warrant, whether from logic, experiment or practice; there can always be grounds for challenging any knowledge claim.[3]

In scientific practice, interpretative flexibility is reduced in processes of accumulation of cognitive and social resources, where

Social Studies of Science Copyright © SAGE Publications (London, Thousand Oaks, CA and New Delhi), Vol. 26 (1996), 611–59

factors like the following play an important role: communication and interaction between research specialties and disciplines, cross-fertilization; interaction between scientific and technological contexts (in this case, information technology and programming); accumulation of organizational, institutional and rhetorical resources; the science-policy context; and the wider scientific and technical culture.[4] By studying the effect of social factors — both internal and external to the scientific community — on the processes of closure of controversies, sociologists have exposed the contingent elements in the production and validation of scientific knowledge.[5]

Specific cognitive objects (such as certain 'crucial' experiments, results or 'proofs') often play an important role in the evolution of scientific controversies. The key move in the controversy analyzed in this paper was the decision by Marvin Minsky and Seymour Papert to replicate the 'Perceptron machine' built by a team led by Frank Rosenblatt, with a view to showing its limitations. As Collins pointed out, replication of this kind is quite unusual in science, and it occurs only when the claim under discussion is particularly important.[6] The 'interpretative flexibility' of Minsky and Papert's results was considerable. Standards of proper experimentation and criteria of competence had not by then been agreed, and experimental work relating to the controversial issues was not equally compelling to all those involved in the debate.

As Trevor Pinch has pointed out, the construction of a disputed cognitive object can be used as the defining point around which differing groups of scientists taking part in a controversy can be identified.[7] These groups emphasize different dimensions of what we can safely assume to be the 'same' object. Disputed cognitive objects can be articulated at different levels. Following Pinch, I will consider two modes of articulation, the 'research-area' mode and the 'official-history' mode.[8] The research-area mode of articulation is used when the disputed object is part of the immediate area of concern and practice of the scientists involved in a controversy, whereas the official-history mode is used in historical accounts of how a particular field evolved. The multi-dimensional character of the object in the research-area mode (the possibility of working on different aspects of it) is lost at the official-history level, where results and proofs are regarded as either valid or invalid. The official-history mode occurs mainly at the informal level of communication in science. Using some concepts and elements from the work of Nicholas Georgescu-Roegen, Richard Whitley, Pierre Bourdieu and others, Pinch showed that the official-history mode of articulation, with its legitimating functions, often plays a very important role in scientific controversies and in the underlying 'battles' for authority in science.

In this paper, I reconstruct the controversy which surrounded a specific 'cognitive object' — namely, certain proofs and arguments which apparently showed that progress in perceptron research was not possible.[9] The structure of the paper and its main arguments are as follows. After a brief presentation of the neural-net approach and its antecedents, I will look at Frank Rosenblatt's Perceptron machine and the controversy which surrounded it. Then I will analyze some of the main technical problems and limitations of early neural nets. I then discuss 'impossibility' proofs and arguments and the process of closure of the controversy, which I reconstruct by linking the developments analyzed previously with the disciplinary, technological and funding contexts. The emergence, institutionalization and legitimation of symbolic AI as a research specialty was the most important factor in this process. Finally, I will examine the process of accumulation and cross-fertilization which has recently brought about the revival of neural nets.

According to the official history of the controversy, in the mid-1960s Minsky and Papert showed that progress in neural nets was not possible, and that this approach had to be abandoned. In this paper I try to show that this official view emerged as a result of the closure of the perceptrons controversy. Before that, things were not so clear at the research-area level. And neural nets were not completely abandoned: a few researchers continued working in this area, but they were displaced from artificial intelligence (AI) to other disciplines.

Collins has used the 'things could have been otherwise' argument in order to show the interpretative flexibility of scientific results. The curious thing about the perceptrons controversy is that things were really otherwise in the recent revival of neural nets, which I will examine at the end of the paper. As neural nets emerged as an accepted specialty, the official history of the controversy was rewritten in order to legitimate the new social and cognitive structures of the AI discipline. The interpretative flexibility of Minsky and Papert's impossibility proofs was reopened.

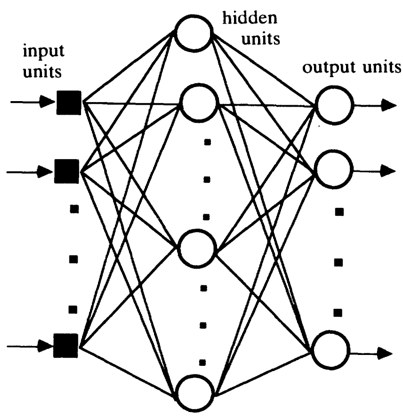
Neural Nets

Neural networks are information-processing systems composed of many interconnected processing units (simplified 'neurons') which interact in a parallel fashion to produce a result or output.[10] They are called 'neural' because, in designing them, researchers are 'inspired' by some (sometimes only a few) simplified features of information processing in the brain. The massively parallel architecture of these systems is remarkably different from that of a conventional (also called 'von Neumann') digital computer. Neural nets are not programmed, but 'trained'. Training a neural net in some classification task involves selecting a statistically representative sample of input/output pairs, and algorithm for adjusting the strengths (or 'weights') of the connections between processing units when the system does not produce the desired outputs. Neural-net training is usually a long (and computationally expensive) process of cycles of input feeding, output observation and weight adjustment.

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

Researchers expect neural nets to have considerable success in tasks not easily programmable so far within the rule-based symbol-processing approach, such as pattern and speech recognition. The learning capabilities of neural nets may be especially important for this type of task. Each unit in a neural-net system performs a simple processing operation which can be divided into three parts: input addition, comparison with a threshold value and, if that value is equalled or surpassed, 'firing' or output activation (see Figure 3, overleaf). Figure 1 shows one of the most popular neural-net architectures — namely, the one formed by strata of units and connections (also called 'multilayer feedforward' net, because activation always spreads in the direction from input to output).

FIGURE 1
Multilayer Network



The architecture of neural nets is radically different from the von Neumann architecture, which is the basis of most existing computers. One of the main characteristics of the von Neumann architecture is the separation between memory and processor. Von Neumann computing consists basically of performing one after another (that is, sequentially) certain transformations upon binary expressions which are stored in the computer memory. These transformations are made according to a list of instructions or rules (the program) which is also stored in the memory. The basic operation of a von Neumann system involves the following steps: localizing an expression in the memory; transferring it to the central processing unit; transforming it; and bringing it back to a different location of the memory.

Von Neumann memory is composed of unrelated, discrete locations, resembling a list of binary expressions (which in AI would stand for symbolic expressions). The only way of finding a certain expression in this list would be to know its exact place. Neural nets work rather differently. Information is not held in discrete locations, but distributed throughout the system's parameters. The 'knowledge' that a neural-net system has at any time of its evolution (that is, the 'state of the system') is given by the activation state of the processing units and the values of the links interconnecting them. In a neural net, linguistic or symbolic expressions do not exist as such, their 'equivalent' (this word has to be used cautiously) being overall patterns of activation emerging from the parallel interaction between many units at the subsymbolic level.

One of the most interesting properties of distributed, associative memories is graceful degradation: in certain circumstances, the net can perform at an acceptable level even if some of its processing units do not work properly (in this, neural computing contrasts sharply with conventional computing and programming). Another important property of neural nets is their ability to recognize whole patterns (or objects), even though only a part of them is presented as an input (or when the image is distorted or incomplete).

Neural-net research is an approach to AI and cognitive science.[11] These related disciplines both aim at building intelligent machines (that is, computer programs and simulations which carry out intelligent or cognitive tasks) and at studying (that is, constructing and testing theories of) perception and cognition using computational methods and tools. AI's emphasis is on building intelligent machines, whereas cognitive science — a highly interdisciplinary field of research — concentrates on understanding cognition.

The origins of AI go back to the cybernetics movement of the 1940s and 1950s. This movement started around the idea that the functioning of many systems, both live and artificial, can be better understood with models based on information processing and transfer, rather than on energy transfer. Researchers aimed at studying the elements that automatic machines and the human nervous system have in common — what they called 'control and communication processes both in the animal and in the machine'. To address this question, an important interdisciplinary effort was made, with contributions from areas like mathematics, formal logic, computer science, psychology, electrical engineering, physiology and neuroscience. The foundations of cybernetics were built and explored by leading scientists, including Alan Turing, Warren McCulloch, Claude Shannon, Norbert Wiener, John von Neumann and Kenneth Craik.

An important aspect of the cybernetics movement was the existence of different approaches to the issue of the relationships between brain (or mental processes) and machine. During the second half of the 1950s, symbol-processing and neural nets were emerging as the two main approaches to both studying cognition computationally (today's cognitive science) and building intelligent machines (today's AI).

The Dartmouth Conference (Hanover, New Hampshire, held as a Summer School in 1956) is usually taken as the starting point of symbolic AI. The emergence phase of this approach ended towards the mid-1960s, when it entered a period of institutionalization and development.[12] Computers had first been used for numerical calculation purposes, but symbolic AI exploited their capability for implementing symbol manipulation. In these systems, symbolic expressions stand for words, propositions and other conceptual entities. The symbol-processing approach is based upon the possibilities that computers offer for storing and processing symbolic expressions. Computers are much better than human beings at storing large quantities of symbolic expressions and processing, manipulating and transforming them in ways sensitive to their logico-syntactical structure. The representational structures contained in a symbolic AI system are manipulated according to certain rules and strategies (programs, algorithms, heuristic rules), and the resulting expression is the solution to a given problem or task. Here information processing occurs at the representational level (its human equivalent would be mental processes), and not at the neurobiological (or brain) level. Symbolic AI systems simulate human mental and cognitive processes by computational (digital, von Neumann) means. Among the most important researchers of early (and contemporary!) symbolic AI were John McCarthy, Allen Newell, Herbert Simon and Marvin Minsky.

But since the early 1950s, in a process which accelerated towards the late 1950s, some researchers had been exploring and developing a different, non-symbolic approach to AI: the so-called neural-net perspective. These scientists and engineers did not seek to model real neural networks as studied by neurophysiology or neurobiology; rather, they were trying to build computational architectures bearing some resemblance to the brain's nets of neurons. These systems were being built employing McCulloch-Pitts artificial or formal 'neurons', connected to each other by links with modifiable links or 'weights' (Donald Hebb's notion of learning by modifying the connections between neurons was foundational in this respect).

The Perceptrons Controversy

Single-layer Machines

In the late 1950s and early 1960s groups from several universities and laboratories carried out research and implementation projects in neural nets. Among the most important projects were those headed by Frank Rosenblatt (Cornell University and Cornell Auronautical Laboratory, CAL), Bernard Widrow (Department of Electrical Engineering, Stanford University) and Charles Rosen (Stanford Research Institute, SRI).

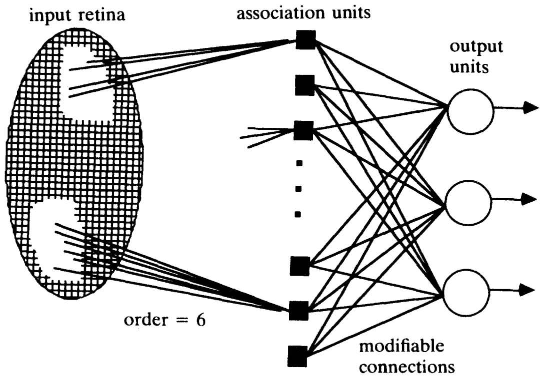
The number of neural-net projects or groups is difficult to quantify. In their critical study of neural nets (analyzed later in this paper), Minsky and Papert alleged that, after Rosenblatt's work, there were perhaps as many as a hundred groups (in an interview conversation this number went up to 'thousands').

> Rosenblatt's (1958) [perceptron] schemes quickly took root, and soon there were perhaps as many as a hundred groups, large and small, experimenting with the model either as a 'learning machine' or in the guise of 'adaptive' or 'self-organizing' networks or 'automatic control' systems.[13]

This issue is an important one in the official-history mode of articulation of the controversy, where it is alleged that Minsky and Papert had to react to stop such a great wave of 'misled' projects. Later I will analyze this issue as a part of the official history. For now it is important to point out that, even though there were not so many projects, neural-net research was one of the main cybernetic approaches to the brain-machine issue, and was taken up very seriously by a significant number of groups and individuals. This can be shown by looking at the scientific meetings of the time, like the 'Mechanisation of Thought Processes' symposium, organized by the British National Physical Laboratory in November 1958, and the 'Self-Organization' conferences held in 1959, 1960 and 1962.[14]

Early researchers made important scientific contributions, especially regarding single-layer neural nets (these were systems with one layer of modifiable connections, although they could have more layers of fixed connections). The most famous machine of this period was Rosenblatt's Perceptron, which is represented in Figure 2. This machine had two layers of connections, but only those from association units to output units had adjustable

FIGURE 2
Perceptron



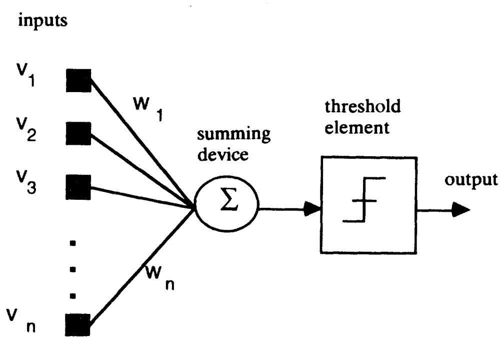
weights. The machine built by Rosenblatt's group at CAL had eight response units, but only three of them are represented in Figure 2. The maximum number of incoming links received by an association unit (or 'order') of this system is six. Later I will show the importance of this issue.

As Figure 3 shows, an output unit fires if the sum of the activation it receives from other units equals or exceeds its threshold value. Note that input activation (v) is multiplied by the values or weights (w) of the connections.

The question of learning was very important in early neural nets (these systems are not 'programmed' in the sense of conventional computers). In order for a perceptron-like net to improve its performance in some classification task, the modifiable connections have to be adjusted according to a rule (or learning algorithm). In 1960, teams led by Frank Rosenblatt, and by Bernard Widrow and Marcian Hoff, developed two very important learning algorithms for single-layer neural nets.[15] Rosenblatt showed that, if a perceptron was physically capable of performing a classification task (that is, if its parameters were capable of embodying that task), then it could be 'taught' that task in a finite number of training cycles.[16] 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.

The perceptron convergence theorem was proved for the simplified perceptron of Figure 4 (representing the adjustable part of the

FIGURE 3
Processing Unit

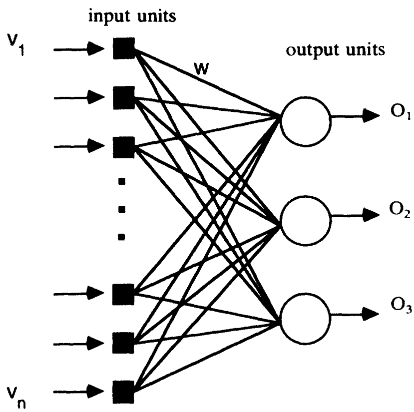


original perceptron after removing the fixed sensory-to-association connections). This algorithm says that, for learning to occur, it is necessary that the perceptron architecture be capable of embodying the desired input/output classification. But proving whether a classification can be carried out by the simplified perceptron of Figure 4 (let alone Rosenblatt's Mark 1 Perceptron, which had a first layer of randomly wired connections) is an NP-complete problem — that is to say, 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 net 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).

The Rhetoric of the Debate

Controversy increased as Rosenblatt's work began to gain notoriety in the late 1950s. Frank Rosenblatt, a psychologist at Cornell University, was the central figure of the early neural-net movement, both from a scientific and from an organizational point of view. He designed and studied the Perceptron, which was implemented at CAL (Buffalo, New York; now the Arvin Calspan Advanced Technology Center), but he was also the charismatic

FIGURE 4
Simplified Perceptron



leader and most enthusiast advocate of neural nets, both within the scientific community and in the wider society.

The Perceptron Project was funded by the US Office of Naval Research (ONR). Rosenblatt and ONR presented it at a press conference held in Washington on 7 July 1958. The statements made by Rosenblatt there, which were widely reported in the mass media, heated the controversy. The following report from *The New York Times* is an example:

> 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.[17]

According to the official history of the controversy, Rosenblatt's 'overclaims' irritated many people in the AI community, including some of its leaders.

> 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, 'a real medicine man. To hear him tell it, the Perceptron was capable of fantastic things . . .'.[18]

Critics accused Rosenblatt of not having respected scientific standards and of having used the media in a partisan way. The following interview quote by Marshall Yovits, who was responsible for the funding of the Perceptron Project at ONR, is interesting in this respect:

> Many of the people at MIT [referring to the symbolic AI leaders] 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 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.[19]

Controversy was rather bitter at times, as scientific arguments, rhetoric and organizational pressure were combined in the process of the debate.

> The campaign was waged by means of personal persuasion by Minsky and Papert and their allies, as well as by limited circulation of an unpublished technical manuscript (which was later de-venomized and, after further refinement and expansion, published in 1969 as the book *Perceptrons*).[20]

The interview quote below, by Charles Rosen (from the SRI group, one of the most important neural-net centres of the time), is another indicator of the tenseness of the debate:

> 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.[21]

Sociologists have shown that rhetoric is an inherent element of discourse and practice in scientific controversies.[22] And, of course, scientists use rhetoric when they present and justify their projects outside the scientific community, as was the case with Rosenblatt. Because of the nature of AI, rhetoric has always been particularly controversial in this discipline.[23] The so-called 'Dreyfus affaire' is one of the most interesting examples, although there are many others. Hubert Dreyfus, a professor of philosophy at the University of California, Berkeley, carefully studied the predictions made by symbolic AI researchers in the 1950s and 1960s, and compared them with the results which were really obtained. The rhetoric studied by Dreyfus included symbolic AI leaders Allen Newell and Herbert Simon's famous 1957 claims that, within ten years, computers would win the world chess championship, compose aesthetically valuable music, discover and prove an important unknown mathematical theorem, and that most psychological theories would take the form of computer programs.

> There are now in the world machines that think, that learn and that create. Moreover, their ability to do these things is going to increase rapidly until — in a visible future — the range of problems they can handle will be coextensive with the range to which the human mind has been applied.[24]

In 1965, Dreyfus wrote a much circulated mimeograph paper which, in 1972, became the basis of his famous *What Computers Can't Do* book.[25] Dreyfus criticized some of symbolic AI's claims from a philosophical point of view. Basically, he argued that the digital, formalized and rule-governed nature of AI was inadequate to model truly human intelligence (with its fuzzy, intuitive, phenomenological and gestaltic aspects). Dreyfus' work provoked a strong reaction from the symbolic AI community, and some interesting and heated debates followed.[26]

In the perceptrons controversy, the contending views were often represented by Rosenblatt and Minsky. They were not only the leaders or spokesmen of the contending positions, but also two of the most important members of the 'core set' of the controversy.[27] Their famous confrontations have been reported in historical accounts of AI.

> 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.[28]

Problems and Limitations of Early Neural Nets

Rosenblatt was aware of the problems and limitations of his Perceptron machine, and acknowledged them in his papers. The machine could not adequately detect similarities between figures, because it classified objects according to the amount of overlap or intersection in the input retina.[29] Preprocessing (distinguishing the components of an image and the relationships between them) was another related problem. Lacking an adequate preprocessing system, a set of association units had to be dedicated to the recognition of each possible object, and so an excessively large layer of association units was needed.[30] Other limitations were excessive learning time and lack of ability to separate parts in a complex environment (Rosenblatt included here the figure-ground or 'connectedness' problem, later analyzed by Minsky and Papert).[31]

Rosenblatt studied more complex architectures: nets with two layers of association units,[32] 'cross-coupled' nets (which had connections among the units of the same layer), and multilayer nets. He claimed that perceptrons' generalization capabilities improved considerably with these changes,[33] but he admitted that very important problems concerning multilayer and 'cross-coupled' nets remained to be solved. Rosenblatt summarized the limitations of perceptrons in a list of fifteen problems, some of which are reproduced below:

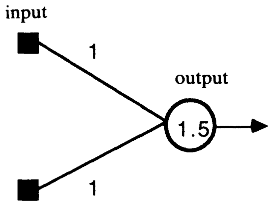
> A number of perceptrons analyzed in the preceding chapters have been analyzed in a purely formal way, yielding equations which are not readily translated into numbers. This is particularly true in the case of the four-layer and cross-coupled systems, where the generality of the equations is reflected 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). Studies of possible figure-segregation (figure-ground) mechanisms. (14) Studies of abstract concept formation, and the recognition of topological or metrical relations. . . .[34]

Rosenblatt's most pessimistic comments were for problems 13 (connectedness) and 14 (recognition of topological relationships and abstract concepts).

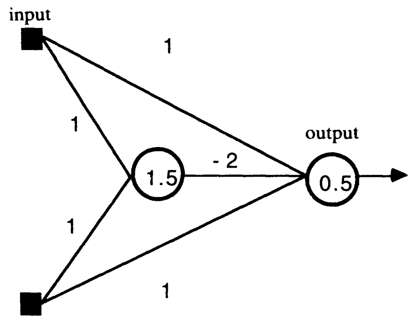
> These two problems [13 and 14] . . . represent the most baffling impediments to the advance of perceptron theory in the direction of abstract thinking and concept formation. The previous questions [from the 1st to the 12th] are all in the nature of 'mopping-up' operations in areas where some degree 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.[35]

FIGURE 5

5.1 'And' Function



5.2 'Exclusive Or' Function



In points 4 and 7, Rosenblatt referred to the difficulties of training multilayer nets, and recognized that this issue 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.[36]

The limitations of single-layer neural nets can be illustrated with a simple example (the simplest one possible). Figure 5.1 shows a simple net composed of two input units and one output unit. It is easy to see that this net can compute the conjunction (or 'and') logical function. The output unit fires only when it receives activation from both input units (only in this case is the sum of input activation bigger than the threshold value, 1.5). But the parameters of the system of Figure 5.1 (the values of the connections and the threshold value) cannot support functions which are not linearly separable, such as exclusive disjunction or 'exclusive-or'.[37] The system should fire when presented input pairs (1, 0) and (0, 1), and should not fire when presented inputs (1, 1) and (0, 0). But if inputs (1, 0) and (0, 1) exceed the threshold value, then input (1, 1) will exceed it too, and the system will fire. As Figure 5.2

shows, an intermediate or 'hidden' unit is necessary in order to realize exclusive-or. This hidden unit would produce strong inhibitory activation (−2) when the (1, 1) input pair is presented to the net.

Early researchers were aware that multilayer systems had much more classification capacity than single-layer ones, but they could not find powerful weight adjustment rules for them.[38]

> 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.[39]

Classifications realized by neural nets can be represented as decision regions in pattern space. Multilayer nets with two layers of hidden units and three layers of modifiable connections (that is, with one more layer of intermediate units than the system of Figure 1) 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).[40] In other words, a multilayer network with two layers of hidden units can realize any input/output classification. Early researchers were aware of the limitations of single-layer systems (some of these will be illustrated in the following section), and there is no doubt that they saw multilayer nets as the way to go. Training multilayer nets was one of the main problems of the early neural-net field.

Early neural nets also had important technological limitations, one of the most important of them being the size of the components. The Perceptron built by Rosenblatt, Charles Wightman and their colleagues at CAL, had only 512 modifiable connections, but it filled a whole laboratory room. Adjustable connections were implemented using motor-driven potentiometers of considerable size, and so implementing a perceptron with thousands of connections using this technology was not practical. Alternative implementations (like the SRI group's magnetic cores, or Widrow's 'memistors') were developed, but this technology was rather limited compared to the emerging von Neumann computer. The advent of the digital computer affected other computer architectures too, like the analog architecture.[41] In fact, certain elements of neural nets (in particular the continuously adjustable weights) and other early cybernetic systems and 'brain models' can be seen as 'analog'.

Digital computers could be used — and indeed started to be used — to simulate neural nets, but the overall philosophy of the neural-net approach, as formulated mainly by Rosenblatt, favoured a brain-style, anti-von Neumann implementational position.

> 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.[42]

> 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.[43]

But even though simulation of neural nets was possible in principle, the association between the digital computer and symbolic AI was much stronger, as I will show later. Before that I will turn to Minsky and Papert's study of the problems of early neural nets. As I pointed out earlier, according to the official-history mode of articulation of the debate these researchers showed that further progress in neural nets was not possible, and after that neural nets were largely abandoned.

The 'Proofs' of the Impossibility of Perceptrons

A Social Service for the AI Community

In the 1950s, several research or problem areas evolved from the cybernetic movement, but none of them had, at that time, yet emerged as a research specialty. Competition became stronger in the late 1950s, as symbolic AI started to emerge as a specialty and neural nets were still attracting a significant amount of human and economic resources. The importance of the problems of early neural nets was not clear. Neural-net researchers maintained that single-layer nets were only the beginning, and that their limitations, important as they were, would be overcome with more complex systems. In the early 1960s, when controversy had reached its highest levels, Marvin Minsky and Seymour Papert, two leading symbolic AI researchers from the prestigious MIT AI group, decided to intervene in the controversy.

> In the middle 1960s 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.[44]

According to the official history, Minsky and Papert were worried by the fact that many researchers were being attracted by neural nets. Their motivating force was (according to this version) to try to stop what for them was an unjustified diversion of resources to an area of dubious scientific and practical value, and to push the balance of AI funding and research towards the symbol-processing side.

> In the late 1950s and early 1960s, after Rosenblatt's work, there was a great wave of neural network research activity. There were maybe thousands of projects. 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].[45]

> 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.[46]

The exaggerated statement about the number of neural-net projects can be understood as part of the official history. Alleging that there were thousands of projects going along such a 'deviant' path justified symbolic AI leaders' strong reaction against neural nets. The social functions of the official history will be analyzed later. Here I will examine Minsky and Papert's technical arguments in some detail. Minsky and Papert's work circulated in the form of drafts and was well known by the mid-1960s, although it was not published as a book until 1969.[47] It is important to note that Minsky and Papert's work had its effect upon the controversy well before the book was published.

In the official-history mode, Minsky and Papert's work is supposed to have shown that further progress in neural nets was not possible, and that therefore this approach lacked scientific or practical value. This is why I will use the term 'impossibility proofs'.[48] However, strictly speaking, Minsky and Papert showed that single-layer nets, defined in a certain way, had some important limitations. On the other hand, they conjectured that progress in multilayer nets would not be possible because of the problem of learning. The key issue (which I try to elucidate below) is that their study was widely seen as a 'knock down' proof of the impossibility of perceptrons (and of neural nets in general).

Minsky had worked in neural nets, but in the early 1950s he abandoned this field to embrace the symbolic approach. It is interesting to note that in the early 1960s he (along with Papert) went back to the neural-net field in order to 'replicate' (so to speak) Rosenblatt's Perceptron, and thus show its limitations. As I have already mentioned, Collins argues that this is rather unusual in science.[49] Normally, one accepts the results coming from an area one is not directly involved with, and the farther away that scientific area is from one's own, the bigger one's certainty about it. Collins pointed out that the crucial and interesting cases are the replication of controversial and important observations, and the core-sets of scientists who are involved in the work. The Perceptron case satisfies this criterion.

Minsky and Papert's work was highly elaborated from a mathematical point of view, and it stands as a very important contribution to neural-net theory. They studied a perceptron similar to the one in Figure 2 (with one output unit instead of three), but they introduced an important restriction regarding the number of connections from input units to association units (the layer of fixed connections in Figure 2). They maintained that the interest of neural computing came from the fact that it was a parallel combination of *local* information, and they suggested that, for this computation to be effective, it had to be 'simple' in some meaningful sense.[50]

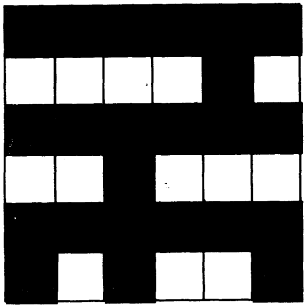
The computation performed by the output unit of their perceptron (a sum of incoming weighted activation in parallel plus a comparison with a threshold) satisfied the proposed criterion. In the case of the association units, Minsky and Papert interpreted their 'simple combination of local information' restriction as implying that each of these units could not receive connections from many input units — that is to say, each association unit could receive connections only from a small part of the input retina. They defined the 'order' of a perceptron as the maximum number of incoming connections received by any association unit (therefore, as I have already mentioned, the order of the perceptron of Figure 2 is 6).

The implications of this 'conjunctive localness' criterion are better understood by looking at the main examples analyzed by Minsky and Papert: 'parity' (saying whether the number of activated inputs in a perceptron retina like the one in Figure 6 is odd or even), and 'connectedness' (the figure-ground problem, consisting of saying whether a set of activated retina points belong to the same object — that is, whether or not they are connected to each other). The problem of parity is related to the exclusive-or function mentioned earlier (in a network with two input units and one output unit, computing parity is equivalent to computing exclusive-or). Minsky and Papert proved that the order required for their single-layer perceptron to compute parity was the whole retina — that is, at least one association unit had to receive connections from all the input units.[51] But 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 could not be satisfied.

The second main problem studied by Minsky and Papert was the 'connectedness' issue. The input pattern appearing in the retina of Figure 6 (the blackened units) is connected. Minsky and Papert proved that the order required for a perceptron to compute the connectedness property also exceeded practical and acceptable limits. This order grew arbitrarily large as the input retina grew in size.[52]

In sum, Minsky and Papert proved that the order required for a perceptron to compute parity and connectedness was not finite; it increased with the size of its input retina. This problem could be seen as equivalent to a conventional computer program having to be rewritten when changing the size of the task.[53]

FIGURE 6



Earlier I showed that early neural-net researchers were well aware of problems like connectedness (especially worrying for object and letter recognition). Nevertheless, in Minsky and Papert's study those problems acquired an 'anomalous' character. Larry Laudan has defined an 'anomalous problem' as a question that both (a) resists solution within a scientific approach, and (b) has an acceptable solution within a competing research tradition;[54] but in controversies, notions like 'resistance to solution' and 'acceptable solution within a competing tradition of research' are evaluated differently by the contending groups. The anomalous character of a problem increases if researchers *agree*, to compare the solution (or the lack of solution) given by a tradition of research with the solution given by a competing one. One important move in Minsky and Papert's rhetoric was to claim that problems such as parity or connectedness could easily be solved using conventional algorithms in serial computers.[55]

> The predicate '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.[56]

> 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.[57]

By emphasizing that parity and connectedness could easily be realized by conventional algorithms in von Neumann computers, Minsky and Papert were linking their critical position about neural

FIGURE 7

nets with two very important factors that would later become closure factors in the controversy: symbolic AI and the digital computer.

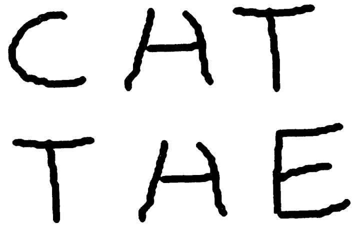
Interpretative Flexibility

But the importance of problems like parity and connectedness was not so clear for neural-net researchers. They compared neural nets not with conventional computers, but with humans. Consider Figure 7.[58] It is not immediately obvious whether the black figure is connected or not. Now look at the white background as a figure. White is connected, and black is not. But this is not obvious the first time one looks at the objects. A conscious, sequential process is necessary in order to determine the connectedness of these figures.

In the research-area mode of articulation, the importance of these problems — and their alleged anomalous character — was open to interpretative flexibility. Neural-net researchers claimed 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 them either.[59] The following quote by David Block, a mathematician from Cornell University who was a colleague of Rosenblatt in the Perceptron project, is an example of this.

> 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

FIGURE 8



Source: H.M. Collins, *Artificial Experts: Social Knowledge and Intelligent Machines* (Cambridge, MA: MIT Press, 1990), 32.

> number of dots in a newspaper photograph *even* or *odd*?), nor connectedness (on the cover of Minsky and Papert's 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.[60]

The relative importance of computing connectedness in certain circumstances can be shown by a letter recognition example. Sometimes connectedness is not a dominant feature of humans' visual environment. The second letter of the two words appearing in Figure 8 is the same (something between 'A' and 'H').[61] One of the appealing properties of neural nets was that, due to previously learnt associations, they would be capable of recognizing whole patterns (in this case the 'A' of 'cat') even though only a part of them (the unconnected second symbol of the first word of the figure) was presented as the input. One of the strong points of neural nets is that, in certain circumstances, they can continue to see the same pattern even when bits are removed that change the figure from connected to disconnected (just like humans!).

Neural-net researchers concentrated on the positive properties of the single-layer perceptron (for example, 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 (systems with more than one layer of adjustable connections, with connections among the units of the same layer, with backward connections, and so on) was needed in order to overcome its limitations. They were asking for time and funding to carry out that research. The issue was, of course, whether their arguments, claims and rhetoric were strong enough to contest Minsky and Papert's criticism.

> 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.[62]

The opinions of other neural-net researchers of the time were similar. For example, Widrow complained that Minsky and Papert had defined the perceptron so narrowly that they could prove that neural nets could do nothing, and he emphasized that his group was working on networks much more complex than the single-layer one.

> 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.[63]

For those actually involved in neural-network research, Minsky and Papert's proofs were (in Widrow's words) 'pretty much irrelevant'. In the research-area mode of articulation, the disputed cognitive objects (Minsky and Papert's 'proofs' and arguments) did not have the static (all-or-none, either valid or invalid) character that is attributed to them in the official-history mode. And, as in Pinch's case study of von Neumann's proof against Bohm in quantum physics, after the perceptrons controversy was closed most people used Minsky and Papert's proofs against neural nets without ever going into them.[64]

As in Pinch's case, the authority of Minsky and Papert's proofs can be linked to the importance of the axiomatic or 'arithmetic ideal' in science,[65] although in this case this ideal should be applied not only to those specific disputed objects but also to the more general differences between the symbolic and neural-net approaches. Symbolic AI is based on the capabilities of the computer for manipulating symbolic expressions in ways sensitive to their logico-syntactical — and therefore discrete — structure. Although the question of proving what a computer program can do is by no means trivial,[66] symbolic AI was much closer to the arithmetic (and rationalist) ideal than the subsymbolic, environment-driven, trained (not programmed) neural-net approach (which was closer to self-organizing, cybernetic systems).[67]

So far I have analyzed Minsky and Papert's proofs about single-layer perceptrons. But what about multilayer nets? The question of learning in multilayer nets had been on neural-net researchers' agenda since the late 1950s, and was widely seen by them as a critical issue. According to the official history of the debate, Minsky and Papert showed that progress in neural nets as a whole (not just in single-layer systems) was not possible. But what Minsky and Papert actually said (in the formal literature) was much less than that.

> 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.[68]

By what process was this conjecture interpreted as showing that further progress in multilayer neural nets was not possible?

Whereas neural-net researchers were asking for time and money for studying more complex systems and trying to solve the problems they had, critics favouring the symbolic perspective claimed that, because of the limits of single-layer systems and the lack of successful learning rules for multilayer systems, progress in neural nets was not possible. We must now analyze the process of closure of this debate — the process through which interpretative flexibility was reduced, and controversy closed. In other words, the question now is to explain the emergence of the official-history view and its social functions.

Closure of the Controversy

Paul Edwards has recently pointed out two important aspects of the emergence of symbolic AI: on the one hand, symbolic researchers' involvement in the early 'mundane practice' (as he puts it) of (von Neumann) computer programming and software development and, on the other, ARPA's institutional support to symbolic AI, mainly through the Information Processing Techniques Office (IPTO, directed by Joseph C.R. Licklider).[69]

The development of high-level computer languages and time-sharing systems was especially important for symbolic AI. As the first computers became commercially available in the 1950s, programmers started to develop compiler and high-level languages in order to simplify the program coding tasks that until then had been done in binary machine language (which was extremely difficult to use and debug). Exploiting the capabilities of the first digital computers required the development of languages which would translate English-like commands and instructions into machine language. As the first high-level programming languages became available in the late 1950s, researchers started to think of computers as manipulators not just of numbers but also of symbolic expressions. Symbolic AI researchers developed programming languages especially suitable for symbol manipulation (such as Newell and Simon's IPL, and McCarthy's LISP).

Symbolic AI programs consumed vast quantities of memory and machine time and, due to the scarcity of the computing resources then available, there was strong competition for computing time. In the late 1950s, computers were 'batch processors' (while a program was running, the machine could do nothing else). Input/output devices were much slower than the central processing unit (CPU), so the CPU was idle most of the time. Edwards describes this situation as follows:

> Programs usually had to be run many times before all errors were found and fixed. Since the debugging process was slower than CPU or input/output times

> by yet further orders of magnitude, after receiving their output and fixing their programs, programmers would have to wait, frustrated, in a queue until the machine was again free. . . . The . . . small number of available computers (especially in universities) meant intense competition for computer time.[70]

Around 1958, John McCarthy developed the idea of CPU time-sharing; he wanted to provide symbolic AI researchers with the possibility of working with LISP interactively from their terminals, without having to deal with the 'priesthood' of computer operators. Working with the computer interactively created the possibility of on-line debugging and fixing of programs while these were running. Thus the effects of each change became instantly visible to the terminal user (the AI researcher).

As Edwards points out, this connection between AI and time-sharing led to the second of the mentioned issues: ARPA's strong support of symbolic AI, mainly through Licklider's IPTO. ARPA's backing of interactive computing, time-sharing systems connected symbolic AI with military projects for human–machine interaction in electronically mediated systems of 'command and control' and 'decision support'. Along with time-sharing, symbolic researchers received strong funding for their scientific objectives of high-level programming, cognitive simulation, heuristics, and the like.

Supported by ARPA funding, the initial leading core of symbolic AI — a reduced group of researchers and their students, working at a few prestigious centres such as MIT (Minsky's group), Carnegie-Mellon University, Stanford University (McCarthy's group), and SRI — had a privileged access to economic and (the then so scarce) computing resources, and consolidated their professional and organizational network.[71] ARPA's policy favoured resource concentration at a few centres of excellence, and selection of projects was based neither on peer review, nor on equalizing principles for research money distribution, but on the agency's own judgement about the best researchers working on the best projects from the point of view of the agency's military goals.[72]

At the same time that it was backing symbolic AI explicitly, ARPA decided — also in an explicit manner — not to fund neural-net research. Both neural-net and symbolic AI researchers were well aware of this, and there is no doubt that this had an impact upon the perceptrons debate. Controversy went beyond the limits of the scientific community, and reached the US Government agencies that were funding AI — mainly ONR and, above all, ARPA. Marvin Denicoff, who worked at ONR in the early 1960s and was also well informed about ARPA's involvement in AI (both agencies collaborated in some respects) told me about this:

> At that time [in the 1960s], the Office of Naval Research had funds at the level of $40K or $50K. ARPA was able to fund hundreds of thousands, or even millions. Rosenblatt never attracted that kind of money, because he wasn't offering a large pay-off. By pay-off I mean not in the scientific sense, but in the application sense, world problem solving. Again, his work was much more, I would say, traditional science. The Office of Naval Research never gave him the kind of money that he really required, and he was not successful in getting the money from the Science Foundation or from ARPA. One can draw the conclusion that if he had had the money he would have made even greater progress. That's too easy an answer, because it doesn't always follow that large amounts of money make the difference. Well before the Minsky and Papert book came, Rosenblatt was not successful in attracting more money, that I know for a fact.[73]

Jon Guice has studied the role of ARPA and the MIT-area defence and research community in the process of closure of the perceptrons controversy. He has documented in detail ARPA's decision to concentrate its IPTO funding resources on the symbolic AI centres from the early 1960s (Minsky's MIT group) and mid-1960s (Stanford, CMU and other smaller institutions), at the same time as it explicitly rejected applications to fund neural-net research.[74] This decision by ARPA was a very important factor in the legitimation of symbolic AI and in the closure of the perceptrons controversy. Guice has also pointed out the importance of an unconventional, satirical paper entitled *Artificial Intelligentsia*, written by consultant Louis Fein in 1963.[75]

Fein asks the reader to imagine that a Federal agency has sent a request to bid on research and development work in AI to four companies. The author then includes the request to bid, the four companies' replies, and an evaluation of the proposals by an external technical expert who advises the agency. Pseudonyms are used for the agency (Bright Field), bidding companies (Optimystica; Dandylines Enterprises; Search Limited, formerly Search Unlimited; and Calculated Risks, Inc.) and evaluator (J.R. 'Bubbles' Piercer, from Pessimyths, Inc., a consulting outfit) The bidding companies represent different AI perspectives and research groups: self-organization (Optimystica), neural nets (Dandylines, which could refer to the SRI neural-net group, perhaps associated with other groups; it is interesting to note that this company sees its work as a continuation of that of Rosenblatt's group), symbolic AI (Search Limited, formerly Search Unlimited, which could well refer to Minsky's MIT group) and probabilistic and statistical pattern recognition, which can be seen as related to neural nets (Calculated Risks, Inc.). 'Bright Field' could be ARPA, and J.R. 'Bubbles' Piercer could be Licklider (who was actually ARPA's IPTO Director from 1962 to 1964).

'Bubbles' Piercer's report contains some interesting points. First, he criticizes Bright Field for failing to ask certain companies to bid (apparently referring to certain neural-net and cybernetics companies). Second, he criticizes the agency's overambitious AI goals, and maintains that AI is in a research phase, far from development and production (he is in favour of AI as an aid to human intelligence, rather than a replacement of it). He points out that AI (including symbolic AI) has made many promises but so far it has failed to deliver. Finally, although he vaguely recommends some support for Calculated Risks and Dandylines (in particular for studying learning and storage capacity in multilayer nets), he ends by making a strong recommendation to support Search Limited. As Guice points out, of these three perspectives of research only the third one (symbolic AI, starting with Minsky's group) was actually funded by ARPA's IPTO.

The (unusual) satirical character of this paper makes it difficult to evaluate, but it can be taken as a (humorous) account of the competition for ARPA funding in the early 1960s between symbolic and other AI approaches (neural nets, and related approaches like probabilistic pattern recognition and cybernetics). ARPA's decision to back symbol-processing and to reject neural nets was a very important closure factor in the perceptrons controversy. It is important to note that, for ARPA, symbolic and heuristic systems were the way to go not only for 'data interpretation and decision-making in command and control' in general, but also for the central areas of interest of neural-network researchers (that is, visual pattern recognition, as applied for example to the interpretation of satellite photographs).

The process of emergence and institutionalization of symbolic AI as a scientific specialty was almost completed by the mid-1960s. By then this approach had accumulated an important stock of scientific contributions.[76] At that time the perceptrons controversy was approaching closure. From the three main early neural-net projects, only Rosenblatt continued his work in perceptrons. Widrow's Stanford University group went into telecommunications engineering applications (where they employed successfully some of their neural-net techniques), and the SRI group started an important mobile robot project within symbolic AI. Later, Rosenblatt's early death in 1971 in a sailing boat accident would leave the neural-net field without its most charismatic leader and advocate.

According to the official history of the controversy, after Minsky and Papert's study, the neural-net approach was rejected and abandoned. Papert himself recognized the existence of 'universalistic' (all-or-none) attitudes.

> Its universalism made it almost inevitable for AI to appropriate our work as a proof that neural nets were universally bad. . . . In fact, more than half of our book is devoted to 'properceptron' findings about some very surprising and hitherto unknown things that perceptrons can do. But in a [scientific] culture set up for global judgement of mechanisms, being understood can be a fate as bad as death.[77]

Papert recognized the existence of a 'global judgement' (against neural nets) in the closure of the perceptrons controversy, and complained that his book with Minsky was interpreted in that sense.

According to the official history, Minsky and Papert replied to Rosenblatt's overclaiming and showed that progress in neural nets was not possible — and after that this field was largely abandoned. But if, as I have shown here, Minsky and Papert did not quite show that, and if (as I will point out soon) neural nets were not completely abandoned, what was the role of the official history? It is my view that its role can only have been the legitimation of the emergence and institutionalization of the symbolic approach, which came to be seen as the 'right' approach to AI, and as occupying the whole AI discipline. In the 1970s, symbolic AI's leading researchers used the 'we are the only AI paradigm' argument in their rhetoric, as can be seen in this quote from a seminal paper by Newell and Simon:

> The principal body of evidence for the symbolic hypothesis that we have not considered [so far in this paper] is negative evidence: the absence of specific competing hypotheses as to how intelligent activity might be accomplished — whether by man or by machine.[78]

Therefore the closure of the perceptrons controversy in the mid-

1960s could well be the 'marker event' that Newell was looking for in his account of the emergence of symbolic AI.

> 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 to 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.[79]

The official history of the debate legitimated the authority structure which was emerging in AI, and was used by the élite of the symbolic approach as a defence strategy against heterodox and 'deviant' interpretations and approaches.

The official history conveniently exaggerates the phenomenon of the abandonment of neural nets. Although neural nets were largely rejected *as an approach to AI*, throughout the 1970s, all over the world, some (not many) researchers — most of them belonging to a younger generation — continued working on neural nets and related topics outside the AI field, in neuroscience and psychology-oriented areas. As the Lighthill report for the UK Science Research Council on the state of AI in the early 1970s shows, neural-network-like research remained somewhat stronger in Europe than in the United States.[80] Researchers who worked in neural nets (in topics such as unsupervised learning and associative memory) in the 1970s include Christoph von der Malsburg, David Willshaw, Teuvo Kohonen, Geoffrey Hinton and Igor Aleksander in Europe; Michael Arbib, Stephen Grossberg, James Anderson, Jack Cowan and Leon Cooper in the United States; and Kunihiko Fukushima and Shun-ichi Amari in Japan.[81]

Therefore the (inaccurate) view of the 'abandonment of neural nets' can be seen as legitimating the emergence of symbolic AI, rather than as an exact description of the result of the perceptrons controversy. After the closure of the controversy, neural-net activity decreased significantly and was displaced to areas outside AI (it was considered 'deviant' within AI) but, contrary to the official view, it did not completely disappear.

The Revival of Neural Nets

Studies of scientific controversies have shown that, once an interpretation has emerged as dominant after the closure of a controversy, time runs against the 'losers' as the organizational and cognitive structures supporting the winning side develop and institutionalize.[82] As the institutionalization of a new social order (with its resource allocation system and authority structure) advances, it increasingly comes to be seen as the only possibility (as a 'natural' order). This is why, in periods of stability, sociologists employ methodological directives such as Everett Hughes's 'remember that it could have been otherwise', in order to remind themselves of the constructed character of social reality. Collins has employed this idea in the sociology of science.[83]

In this paper, I have tried to show the interpretative flexibility of Minsky and Papert's proofs and arguments about the impossibility of perceptrons. The rejection of neural nets as an approach to AI was a contingent social process, and therefore, in principle, 'things could have been otherwise'. The interesting and curious thing about neural nets is that things were actually otherwise in the middle and late 1980s, two decades after the closure of the perceptrons controversy. Here I can only review briefly some of the main developments which brought about this change.[84]

In the early 1980s, symbolic AI went from institutionalization to a stage of growth, applications and (the beginning of) commercialization.[85] International competition and interest in this specialty increased as the US and UK governments reacted to the announcement by the Japanese Government of its Fifth Generation Project (especially directed to areas like natural language processing and 'knowledge engineering', or knowledge-based systems). The rise of the expert systems application area in the mid-1980s was one of the main developments of this period.[86]

Developments in information technology were a major change affecting AI. In the early 1980s, dramatic decreases in computing costs brought about a 'democratization' in the access to computing resources. As a result, as James Fleck points out, the scope for the strong symbolic AI élite to control the development of the field was weakened, allowing outsiders to move in and pursue their own variants of AI research.[87] Eventually this permitted the use of powerful computing resources to simulate until-then 'deviant' approaches, such as neural nets. On the other hand, since the late 1970s, researchers from a variety of fields in the human sciences (and later in the neurosciences) had started to use the computer as a research tool in an emerging interdisciplinary discipline called 'cognitive science'. Although cognitive science was then based in the symbolic approach, its interdisciplinary character helped bring new perspectives into the computer-mind problem.

In this context, some symbolic AI researchers started to confront the limitations of their models. Expert systems were being applied to a great variety of problems, but symbolic AI was not so successful in areas such as speech recognition, pattern recognition, and common-sense and heterogeneous reasoning. Some researchers started to look at new approaches for studying and modelling these tasks.

The conference organized in June 1979 at La Jolla (California) by neural-net 'veterans' Geoffrey Hinton and James Anderson can be seen as the first contact between researchers who had been working in neural nets throughout the 1970s and researchers coming from the symbolic approach, but looking for ways of solving some of its limitations. The papers presented there were developed and published in 1981 in a book entitled *Parallel Models of Associative Memory*.[88] The topics of the book are a good sample of the perspectives which were being considered: information processing in the brain, connectionist local nets, semantic nets, and associative memory. Other topics which these researchers were looking at include parallelism in vision research (for example, interaction between many local features in the interpretation of an image) and multiple constraint systems.[89] After this, the Parallel Distributed Processing (PDP) group was formed in the University of California-San Diego, headed by psychologists David Rumelhart and James McClelland.

Although neural nets were not directly linked to the neurosciences, increases of activity and interest in the latter in the 1980s contributed to a more favourable context for the former. PDP and other researchers adopted a 'brain-style' style of information processing. They argued that the information processing power of the brain comes from its parallelism. Given the facts that neurons are not too fast (firing frequencies range from a few to a few hundred impulses per second) and that some complex mental behaviour (like recognizing a face) takes 1/10 second, researchers concluded that the brain's information processing power must come from its parallelism.[90]

The advent of parallel computers and supercomputers in the 1980s as an attempt to overcome the speed limitations of sequential computers (separation between memory and central processing unit in a von Neumann computer imposes a sequential, 'one operation at a time' style of computation) added plausibility to 'brain-style' computation. As with the neurosciences, the connection between parallel computers and neural nets was not straightforward in the beginning; many of the most successful neural-net experiments of the mid-1980s were done as simulations in sequential computers. On the other hand, there are many parallel computer architectures, and neural nets are one extreme type (massively parallel).[91] Nevertheless, increases in computer power and speed due to parallelism will undoubtedly favour neural-net research.[92]

The work done by the PDP group (with people like Rumelhart, McClelland, Hinton and Terrence Sejnowski) and by 'veterans' such as Anderson, Grossberg, Kohonen, Willshaw and von der Malsburg started to a attract researchers from other disciplines to the neural-net field.[93] Migration is a common phenomenon when a new area of research is emerging.[94] Researchers coming from overpopulated areas or specialties, or having widely applicable backgrounds such as physics or mathematics, may perceive interesting or non-exploited problems and career opportunities in different, emerging areas.

The case of John Hopfield, a physicist from the California Institute of Technology, was particularly important.[95] Hopfield used a method of the physics of collective phenomena (the Ising model of magnetic material, or 'spin-glass') in order to develop a new neural-net architecture with symmetric connections that could be used as an associative content-addressable memory.[96]

> In physical systems made from a large number of simple elements, interactions among large numbers of elementary components yield collective phenomena

> such as the stable magnetic orientations and domains in a magnetic system. Any physical system whose dynamics in phase space is dominated by a substantial number of locally stable states to which it is attracted can therefore be regarded as a content-addressable memory. The physical system will be a potentially useful memory if, in addition, any prescribed set of states can readily be made the stable states of the system.[97]

Hopfield's model was later developed by Hinton and Sejnowski, two of the most important researchers of the PDP group, into the 'Boltzmann machine' stochastic multilayer net.[98] Hinton and Sejnowski developed a learning algorithm which usually got the best global minima and, although in the beginning it was quite slow, they presented it as a first solution to the problem of learning in multilayer nets.

> In the Boltzmann machine, Hinton and I found a learning algorithm which overcame the conjecture by Minsky and Papert that you couldn't generalize the perceptron learning algorithm to a multilayered architecture.[99]

> A learning algorithm was discovered for the Boltzmann machine that provided the first counterexample to the conjecture by Minsky and Papert that extensions of the perceptron learning rule to multilayered networks were not possible.[100]

Both the Hopfield and the Hinton-Sejnowski cases show that cross-fertilization and communication between neural nets and other scientific fields (physics of collective phenomena; stochastic techniques from statistical mechanics) was very important in the neural-net revival.[101] Different techniques were applied to the study of representation and learning in nonlinear dynamical neural net systems.

After the Boltzmann net, PDP researchers Rumelhart, Hinton and Ronald Williams developed a learning algorithm for multilayer feedforward (that is, perceptron-like) nets, the so-called back-propagation algorithm.[102] This contribution — the most popular of the neural-net revival — triggered a new wave of neural-net research. Figure 1 earlier represents the type of architecture for which Rumelhart and his colleagues developed their technique. The main problem for weight adjustment in multilayer nets is to know the error made by the hidden units, in order to be able to adjust the connections between input units and hidden units (the error made by the output units is the difference between the real output pattern and the desired one). 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 it 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 net, each output unit demands from the hidden units exactly what it needs, and the hidden units try to accommodate the conflicting demands.[103]

A very important difference between the back-propagation net and the perceptron was the introduction of smooth or sigmoid activation functions (in the processing units) instead of the classic discontinuous step functions, so that it became possible to compute error gradients in multilayer feedforward nets (the derivatives of the error with respect to the hidden units' output could be calculated). This small change in the assumptions defining a neural net made possible the study of complex systems with flexible activation surfaces.

> Small reformulations of a problem can greatly change the possibilities of making progress. The change from threshold logic units to sigmoids might not seem like a major reformulation, but by using continuous rather than discontinuous functions, it became possible to generalize the Widrow-Hoff and perceptron learning algorithms to multilayered networks.[104]

It is interesting to note that Paul Werbos developed a technique equivalent to back-propagation in the 1970s, but found resistance to his idea of applying it to neural nets.[105]

In 1986, PDP researchers Rumelhart and McClelland sent a report to DARPA and the National Science Foundation (NSF) asking for funding for neural nets and warning against further neglect of this approach.[106] DARPA's *Neural Network Study*, and its subsequent decision to start support for this approach, were especially significant because of the strong role played by this agency in the development (and legitimation) of symbolic AI. By the end of the 1980s, most US European and Japanese funding agencies had launched programmes in neural nets.

In the process of legitimation of the new neural-net movement of the late 1980s, the PDP researchers confronted the view which had helped legitimate the symbolic approach (and delegitimize neural nets) in the 1960s — namely, the official history of the controversy. Rumelhart and his colleagues claimed that, even though their back-propagation net sometimes got trapped in local (or false) minima, in practice the system led to acceptable solutions in 'virtually every case'. They claimed that they had overcome Minsky and Papert's impossibility proofs and arguments:

> 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.[107]

In a sense, PDP researchers made use of the official history for their own benefit. They were saying something like 'after all, Minsky and Papert did not really show that neural nets were impossible'. They were exploiting the interpretative flexibility of the debate to their own benefit, but it is important to note that they were able to do this within the process of accumulation and cross-fertilization of the middle and late 1980s. Other people who tried to do the same before then (like Werbos with back-propagation, or some neural-net 'veterans' with other systems) failed.[108] The official history of the debate was rewritten in order to legitimate the 'new order' (resource allocation system and authority structure) resulting from the revival of neural nets, and its emergence as an AI research specialty.

Rumelhart and his colleagues' claims reopened the controversy, and Minsky and Papert reacted quickly.

> 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. . . . 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.[109]

As the neural-net revival advanced in the late 1980s, the controversy about the validity and feasibility of neural nets (the old perceptrons controversy) reopened, and there were new episodes of interpretative flexibility.[110] But this time the emergence of neural nets as an AI specialty was unstoppable.[111] Techniques like back-propagation were developed and applied to a wide variety of practical problems in areas such as object and speech recognition.[112]

Debate about the relationships between the symbolic and neural-net approaches continued, but the most negative views about the neural-net field were quickly overcome.[113] As both approaches were compared and developed, the strong and weak points of each of them was being tested in each particular problem. After a first period of quite strong competition between the two approaches, the situation will probably evolve into a more normalized combination of competition and — increasingly — cooperation.

Concluding Summary

In this paper, I have analyzed the controversy which surrounded Rosenblatt's Perceptron Project (and neural nets in general) in the late 1950s and early 1960s. Attention has been focused on a particular cognitive object: Minsky and Papert's proofs and arguments, which were interpreted as showing that further progress in neural nets was not possible and that therefore this approach had to be abandoned. I have distinguished two modes of articulation of this disputed cognitive object: the research-area mode and the official-history mode.[114] I have shown that the official-history mode of articulation played a crucial role in the controversy.

At the research-area level, there was considerable interpretative flexibility about Minsky and Papert's proofs and arguments. Scientists using different research techniques and having different approaches and interests interpreted those results differently. However, as the symbolic AI approach emerged and institutionalized, an official interpretation emerged according to which Minsky and Papert had shown that progress in perceptrons — and in neural nets in general — was not possible. According to this official-history view, neural nets were abandoned in the late 1960s.

The official-history mode of articulation of the debate can be seen as part of the discourse of legitimation of the new AI 'order' (with its resource allocation system and authority structure) which emerged from the institutionalization of the symbolic approach as a research specialty. The symbolic approach was presented as occupying the whole AI field, and the official history of the perceptron debate was used as a defence strategy against 'deviant' claims and approaches (such as neural nets). Some researchers continued working in neural-net-related topics throughout the 1970s, but they were displaced from the AI field.

The interpretative flexibility of the debate is further shown by the revival of neural nets (in different circumstances) in the mid-1980s. In the recent process of emergence and legitimation of neural nets as an AI research specialty, the official history was revised (PDP researchers claimed that 'after all, Minsky and Papert did not really show that progress in neural nets was impossible') as the AI field was being socially and cognitively redefined, and a new resource allocation system and authority structure was developing.

• *NOTES*

The work on which this paper is based was supported by a Basque Government scholarship for doctoral research at the Department of Sociology of the University of Edinburgh (1988–91). I would like to thank the following people for advising, encouraging and helping me throughout my research: Donald MacKenzie (PhD supervisor), James Fleck, Alfonso Molina and David Willshaw (University of Edinburgh); Peter Dayan (University of Toronto); and Jesus Maria Larrazabal and

Jesus Ezquerro (University of the Basque Country). Many thanks are due to Joel Kallich and Sue Jennings, who helped me in my research trip to the United States. Finally, I would like to thank both the people I interviewed and those who sent me information by letter.

1. H.M. Collins (ed.), 'Knowledge and Controversy: Studies of Modern Natural Science', *Social Studies of Science*, Vol. 11, No.1 (February 1981), 3–158; Collins, 'The Place of the "Core Set" in Modern Science: Social Contingency with Methodological Propriety in Science', *History of Science*, Vol. 19 (1981), 6–19; Collins, 'An Empirical Relativist Programme in the Sociology of Scientific Knowledge', in Karin D. Knorr-Cetina and Michael Mulkay (eds), *Science Observed: Perspectives on the Social Study of Science* (London: Sage, 1983), 85–113; Collins, *Changing Order: Replication and Induction in Scientific Practice* (London: Sage, 1985); Bruno Latour, *Science in Action: How to Follow Scientists and Engineers throughout Society* (Milton Keynes, Bucks.: Open University Press, 1987); S. Leigh Star, *Regions of the Mind: Brain Research and the Quest for Scientific Certainty* (Stanford, CA: Stanford University Press, 1989).

2. Collins (1983), op. cit. note 1.

3. This is Donald MacKenzie's formulation: see D. MacKenzie, *Inventing Accuracy: A Historical Sociology of Nuclear Missile Guidance* (Cambridge, MA: MIT Press, 1990), 10.

4. My idea of the processes of accumulation of resources and cross-fertilization is compatible with some parts of Bruno Latour's *Science in Action*, op. cit. note 1. I take this book as a continuation of Collins' and others' controversy studies. Other parts of actor-network theory — which are supposed to be 'philosophically more radical' — are not useful for my case study.

5. As sociologists have argued, since the pioneer work of Barry Barnes and David Bloor, this is not a characteristic of 'bad' or 'ideological' science, but of all science: see, for example, B. Barnes, *Scientific Knowledge and Sociological Theory* (London: Routledge & Kegan Paul, 1974); D. Bloor, *Knowledge and Social Imagery* (London: Routledge & Kegan Paul, 1976).

6. Collins, 'Core Set', op. cit. note 1. I will come to this issue later.

7. T.J. Pinch, 'What Does a Proof Do if it Does Not Prove?: A Study of the Social Conditions and Metaphysical Divisions Leading to David Bohm and John von Neumann Failing to Communicate in Quantum Physics', in Everett Mendelsohn, Peter Weingart and Richard Whitley (eds), *The Social Production of Scientific Knowledge, Sociology of the Sciences Yearbook*, Vol. 1 (Dordrecht: D. Reidel, 1977), 171–215, at 174. In this paper, I use Pinch's case study as a parallel, and I apply many of the categories he uses to my own case study.

8. Pinch, op. cit. note 7, 174–76.

9. For a detailed account of the evolution of neural-network research, see Mikel Olazaran, *A Historical Sociology of Neural Network Research* (unpublished PhD dissertation, Department of Sociology, University of Edinburgh, 1991); Olazaran, 'A Sociological History of the Neural Network Controversy', *Advances in Computers*, Vol. 37 (1993), 335–425. In the present paper, I examine just one key issue of the evolution of neural nets: the emergence and functions of the official history of the perceptrons controversy.

10. Neural networks are also called 'artificial neural networks', 'connectionist networks', 'parallel distributed systems' and 'neural computing systems'.

11. Neural nets also have a very important engineering aspect, directed toward special purpose hardware implementation. Nevertheless, the main developments of the evolution of neural nets have occurred around AI. AI can be seen as the core of a wider, interdisciplinary field called 'cognitive science'. Although cognitive science did not emerge as a differentiated discipline until the late 1970s, many of its main problems were studied under different headings earlier.

12. For the history of symbolic AI, see: James Fleck, *The Structure and Development of Artificial Intelligence: A Case Study in the Sociology of Science* (unpublished MSc dissertation, University of Manchester, 1978); Pamela McCorduck, *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence* (New York: W.H. Freeman, 1979); Fleck, 'Development and Establishment in Artificial Intelligence', in Norbert Elias, Herminio Martins and Richard Whitley (eds), *Scientific Establishments and Hierarchies, Sociology of the Sciences Yearbook*, Vol. 6 (Dordrecht: D. Reidel, 1982), 169–217; Fleck, 'Postscript: The Commercialisation of Artificial Intelligence', in Brian P. Bloomfield (ed.), *The Question of AI* (London: Croom-Helm, 1987), 149–64; Paul N. Edwards, *The Closed World: Computers and the Politics of Discourse in Cold War America* (Cambridge, MA: MIT Press, *Inside Technology* series, 1995, forthcoming).

13. M.L. Minsky and S.A. Papert, *Perceptrons: An Introduction to Computational Geometry* (Cambridge, MA: MIT Press, 1969), 19. Minsky and Papert refer to F. Rosenblatt, 'The Perceptron: a Probabilistic Model for Information Storage and Organization in the Brain', *Psychological Review*, Vol. 65 (1958), 386–408.

14. National Physical Laboratory (NPL), *Mechanisation of Thought Processes*, Vols I and II (London: Her Majesty's Stationery Office, 1959); Marshall C. Yovits and S. Cameron (eds), *Self-organizing Systems: Proceedings of an Interdisciplinary Conference*, Chicago, IL, 5–6 May 1959 (New York: Pergamon Press, 1960); H. von Foerster and G.W. Zopf (eds), *Illinois Symposium on Principles of Self-organization*, University of Illinois, Urbana, IL, 1960 (New York: Pergamon Press, 1962); Yovits, G.T. Jacobi and G.D. Goldstein (eds), *Self-organizing Systems 1962* (Washington, DC: Spartan, 1962). In the 'Mechanisation of Thought Processes' conference there were contributions from approaches including symbolic AI (M. Minsky and J. McCarthy), 'cybernetics' (Donald M. MacKay, W. Ross Ashby), pattern recognition (Oliver G. Selfridge, A.M. Uttley, Warren S. McCulloch and Wilf K. Taylor) and neural networks (F. Rosenblatt). In the 1962 conference on self-organization, there were contributions from perspectives including neural modelling (Leon D. Harmon), brain theory/neural networks (W.S. McCulloch, Michael A. Arbib, Jack D. Cowan), neural networks (F. Rosenblatt, B. Widrow), neural networks/electrophysiological experiments (B.G. Farley), symbolic AI (A. Newell) and 'cybernetics' (D.M. MacKay). Within the cybernetics movement, work was done which was related to neural networks in diverse ways and degrees. Oliver Selfridge's (NPL, op. cit., 511–26) hybrid Pandemonium system is an example; another one is work on pattern recognition in Britain by A.M. Uttley (ibid., 119–47) For early neural-network papers, see also J.A. Anderson and Eduard Rosenfeld (eds), *Neurocomputing: Foundations of Research* (Cambridge, MA: MIT Press, 1988).

15. F. Rosenblatt, *On the Convergence of Reinforcement Procedures in Simple Perceptrons* (Buffalo, NY: Cornell Aeronautical Laboratory Report VG-1196-G-4, 1960); Rosenblatt, *Principles of Neurodynamics* (New York: Spartan, 1962); B.

Widrow and M.E. Hoff, 'Adaptive Switching Circuits', *1960 IRE WESCON Convention Record* (New York: IRE, 1960), 96–104.

16. Rosenblatt, *Principles of Neurodynamics*, op. cit. note 15, 111.

17. 'New Navy Device Learns by Doing', *The New York Times* (8 July 1958), 25:2. For similar statements, see: 'Electronic "Brain" Teaches Itself', *The New York Times* (13 July 1958), at iv 9:6; 'Rival', *The New Yorker* (6 December 1958), 44–45.

18. McCorduck, op. cit. note 12, 87. For a list of people irritated by Rosenblatt, see ibid., 88.

19. Yovits, interview, 28 November 1989.

20. Robert Hecht-Nielsen, *Neurocomputing* (Reading, MA: Addison-Wesley, 1990), 16–17. Hecht-Nielsen refers to Minsky & Papert, op. cit. note 13.

21. Rosen, interview, 10 November 1989.

22. See Collins (1983), op.cit. note 1, 49; Latour, op. cit. note 1; Star, op. cit. note 1.

23. This is due in part to the fact that AI affects social discourses about the similarities and differences between human beings and machines: see J. Fleck, 'Artificial Intelligence and Industrial Robots: An Automatic End for Utopian Thought', in E. Mendelsohn and Helga Nowotny (eds), *Nineteen Eighty-Four: Science between Utopia and Dystopia, Sociology of the Sciences Yearbook*, Vol. 8 (Dordrecht: D. Reidel, 1984), 189–231.

24. Cited in Hubert L. Dreyfus, *What Computers Can't Do: The Limits of Artificial Intelligence* (New York: Harper Colophon, 2nd edn, 1979), 81–82. See also Edwards, op. cit. note 12, 315 (draft publication).

25. Dreyfus, op. cit. note 24.

26. Some parts of Dreyfus's work are not far from the kind of contributions that the sociology of knowledge could make to AI and cognitive science. It is interesting to note that, although Dreyfus's work provoked a strong critical reaction from the AI leaders in the 1960s (see McCorduck, op. cit. note 12, Chapter 9), recently someone as qualified as Minsky has implicitly recognized that such philosophical research can make positive contributions to AI. See the following quote from Minsky's opening talk at the 1988 IEEE International Conference on Neural Networks, in the heat of the neural-net revival: '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', from Randolph K. Zeitvogel, 'ICNN Reviewed', *Synapse Connection* (now *Neural Technology Update*), Vol. 2–8 (1988), 10–11. For a recent discussion about AI from the perspective of the sociology of knowledge, see H.M. Collins, *Artificial Experts: Social Knowledge and Intelligent Machines* (Cambridge, MA: MIT Press, 1990).

27. The use of the 'core set' concept in controversy studies is due to Collins, 'Core Set' & *Changing Order*, op. cit. note 1.

28. McCorduck, op. cit. note 12, 88.

29. Rosenblatt, op. cit. note 13, reprinted in Anderson & Rosenfeld (eds), op. cit. note 14, 92–114, at 96; Rosenblatt, *Principles of Neurodynamics*, op. cit. note 15, 67–70; F. Rosenblatt, 'Strategic Approaches to the Study of Brain Models', in von Foerster & Zopf (eds), op. cit. note 13, 385–96, at 390–91.

30. Rosenblatt, *Principles of Neurodynamics*, op. cit. note 15, 306.

31. Ibid., 309–10.

32. In a perceptron with two layers of association units, the units of the first association layer which responded to similar features in different positions would all activate the same unit in the second association layer, and in this way a feature in different positions could be recognized as the same.

33. Rosenblatt, *Principles of Neurodynamics*, op.cit. note 15, 576.

34. Ibid., 577–79. 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.

35. Rosenblatt, *Principles of Neurodynamics*, op. cit. note 15, 580–81.

36. Ibid., 579–80.

37. For two propositions, the truth value of the exclusive disjunction function is as follows:

p	q	ex-or
1	1	0
1	0	1
0	1	1
0	0	0

38. There are other important issues in neural-network research, such as the number and type of processing units needed in order to carry out a task, the number and type of layers of units, and the number of weight adjustment cycles required, but the lack of a learning algorithm is an underlying, fundamental problem.

39. J.K. Hawkins, 'Self-Organizing Systems: A Review and Commentary', *Proceedings of the Institute of Radio Engineers* (IRE), Vol. 49 (1961), 31–48, quote at 45–47.

40. Richard P. Lippmann, 'An Introduction to Computing with Neural Nets', *IEEE ASSP Magazine*, Vol. 4 (1987), 4–22, at 15–18; DARPA, *Darpa Neural Network Study* (Fairfax, VA: Armed Forces Communications and Electronics Association [AFCEA] International Press, 1988), 78–80.

41. The history of the analog computer (which was used for solving differential equations) goes back to the 1930s, but its golden age was the 1950s and early 1960s. Voltage precision problems in analog computers contrasted with the advances in accuracy, speed, memory capacity, miniaturization and programming of digital computers. By the mid-1960s, digital technology was the choice for most computer users: see Time-Life Books, *Alternative Computers* (Alexandria, VA: Time-Life Books, 1989), 26, 27 & 39.

42. Rosenblatt, *Principles of Neurodynamics*, op. cit. note 15, 10.

43. F. Rosenblatt, 'Two Theorems of Statistical Separability in the Perceptron', NPL, op. cit. note 14, Vol. I, 421–56, at 422.

44. Jeremy Bernstein, 'Profiles: AI, Marvin Minsky', *The New Yorker* (14 December 1981), 50–126, quote at 100.

45. Minsky, interview, 25 October 1989.

46. S.A. Papert, 'One AI or Many?', in Stephen R. Graubard (ed.), *The Artificial Intelligence Debate: False Starts, Real Foundations* (Cambridge, MA: MIT Press, 1988), 1–14, quote at 4–5, emphasis in original.

47. Minsky & Papert, op. cit. note 13.

48. Of course, I am not using the term 'proof' in an absolute sense. Here I

relativize this term, as has been done in other case studies: see Pinch, op. cit. note 7.

49. Collins, 'Core Set', op. cit. note 1, 7, 8 & 14.

50. Minsky & Papert, op. cit. note 13, 9.

51. Ibid., Chapter 3.

52. Ibid., 8, 17 & Chapter 5.

53. Igor Aleksander and Helen Morton made this comparison: 'Minsky and Papert's 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': I. Aleksander and H. Morton, *An Introduction to Neural Computing* (London: Chapman & Hall, 1990), 41.

54. Larry Laudan used the term 'anomalous problem': see L. Laudan, *Progress and its Problems: Towards a Theory of Scientific Growth* (Berkeley, CA: University of California Press, 1977), 29. I am, therefore, not using this term in its Kuhnian sense (experimental results which do not fit within the accepted categories of a scientific paradigm).

55. Aleksander and Morton described some simple algorithms for computing parity and connectedness in input retinas like the one in Figure 6: '(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 & Morton, op. cit. note 53, 39–40).

56. Minsky & Papert, op. cit. note 13, 72.

57. Ibid., 227.

58. This figure is inspired by the drawing appearing on the front page of Minsky & Papert, op. cit. note 13.

59. The relationship between human information processing (whether at the mind or brain level) and machine information processing has been a constant and especially important rhetorical resource throughout the history of AI, because of the prominent role played in this discipline by the 'computer metaphor'.

60. H. David Block, 'A Review of *Perceptrons*', *Information and Control*, Vol. 17 (1970), 510–22, quote at 517, emphasis in original. Block refers to Minsky & Papert, op. cit. note 13.

61. This example is taken from Collins' discussion about the distinction between behaviour and action. In the case of this pseudoletter, two different intentions (the 'A' of 'cat' and the 'H' of 'the') were executed by the same behaviour (the same movements of the hand): see Collins, op. cit. note 26, 32.

62. Block, op. cit. note 60, 513–14, emphasis in original.

63. Widrow, interview, 13 November 1989.

64. Pinch, op. cit. note 7, 189 & 205–06. The opinions of some physicists interviewed by Pinch have strong similarities with Widrow's view; for example: 'Well, I suppose that [most physicists] regard von Neumann's book as a perfectly adequate formal treatment for pedants, people who like that sort of thing. They wouldn't read it themselves but they're glad someone has done all that hard work!' (ibid., 205).

65. The concept of an 'arithmetic ideal' goes back to Georgescu-Roegen's concept of 'arithmomorphism', and has been developed in the sociology of science mainly by Richard Whitley: see R. Whitley, 'Changes in the Social and Intellectual Organization of the Sciences: Professionalization and the Arithmetic Ideal', in Mendelsohn, Weingart & Whitley (eds), op. cit. note 7, 143–69.

66. For a recent contribution to the issue of proofs of computer-system correctness, see D. MacKenzie, 'Negotiating Arithmetic, Constructing Proof: The Sociology of Mathematics and Information Technology', *Social Studies of Science*, Vol. 23 (1993), 37–65.

67. Minsky and Papert exploited this point in their own rhetoric, and spoke about 'mystique' and 'a certain flourish of romanticism' surrounding 'loosely organized and distributed neural network machines': see Minsky & Papert, op. cit. note 13, 4 & 18–19. Proving what a symbolic AI system can do is a much more open and 'softer' question than in conventional computer science (typically, symbolic AI systems are based on heuristic, rather than algorithmic, searches).

68. Minsky & Papert, op. cit. note 13, 231–32.

69. See Edwards, op. cit. note 12, Chapter 8. 'ARPA' (the Advanced Research Projects Agency of the US Department of Defense, nowadays known as 'DARPA') was created by the Eisenhower administration in 1958, as a reaction to the Sputnik launch, and in a period of unprecedented growth in funding for basic research in the United States (late 1950s and early 1960s). As Jon Guice points out (see below: op. cit. note 74), ARPA's early programmes were linked to Cold War policy concerns, including nuclear test detection, and space and missile technologies. From the late 1960s onwards, ARPA's programmes diversified. My interest is in ARPA's involvement in the development of AI, as described in the text.

70. Edwards, op. cit. note 12, Chapter 8, 317 (draft pagination).

71. Fleck (1982), op. cit. note 12. Fleck points out that up until the mid-1970s, symbolic AI research was backed almost exclusively by ARPA funding (ibid., 181).

72. McCorduck, op. cit. note 12, 110.

73. Denicoff, interview, 29 November 1989. Denicoff refers to Minsky & Papert, op. cit. note 13.

74. Jon Guice, 'Lord ARPA and the Battle of Perceptrons: Controversy and the State in Intelligent Computing, 1958–69', draft paper under consideration by *Social Studies of Science*; Guice, *Designing the Future: The US Advanced Research Projects Agency* (La Jolla, CA: Department of Sociology, University of California at San Diego, PhD in progress, working title).

75. Louis Fein, 'The Artificial Intelligentsia', *Wescon Technical Papers*, Vol. 7 (11.1, Part 7) (1963), 1–7. This paper was later published with some minor changes, as: L. Fein, 'The Artificial Intelligentsia', *IEEE Spectrum* (February 1964), 74–87.

76. Including, among others, the following: McCarthy's LISP and Newell's ILP programming languages; programs which combined algorithmic and heuristic methods, such as Newell, Simon and Shaw's General Problem Solvers; the Logic

Theorist theorem-proving program; chess playing systems; Minsky's work on heuristic problem solving; the Stanford DENDRAL system for the analysis of the molecular structure of unknown composites; and the MIT MACSYMA system for mathematical problem solving.

77. Papert, op. cit. note 46, 7–8.

78. A. Newell and H.A. Simon, 'Computer Science as Empirical Enquiry: Symbols and Search', *Communications of the Association for Computing Machinery*, Vol. 19 (1976), 113–26, reprinted in John Haugeland (ed.), *Mind Design* (Cambridge, MA: MIT Press, 1981), 35–66, quotation from reprinted version, at 50.

79. A. Newell, 'Intellectual Issues in the History of Artificial Intelligence', in Fritz Machlup and Una Mansfield (eds), *The Study of Information: Interdisciplinary Messages* (New York: John Wiley & Sons, 1983), 187–227, at 201–02, emphasis added.

80. James Lighthill, *Artificial Intelligence* (London: Science Research Council, 1973). The three main areas of AI research studied by Lighthill were (neuroscience and psychology-oriented) computer-based central nervous system research in man and animals, symbolic AI/advanced automation, and robotics. Lighthill concluded that success of work under the category 'computer-based central nervous system research' (neural network-like research) would depend on its close relationships with psychology and neurobiology, in the same way as work on advanced automation/symbolic AI would depend on its close association with its application area (engineering): see ibid., 19–21. He also maintained that robotics research should integrate in those areas.

81. Some representative neural-net contributions of the 1970s were reprinted in Anderson & Rosenfeld, op. cit. note 14. Vision researcher David Marr worked in neural net-like topics until the early 1970s, when he switched to the symbolic approach.

82. See, for example, Bill Harvey, 'Plausibility and the Evaluation of Knowledge: A Case-Study of Experimental Quantum Mechanics', *Social Studies of Science*, Vol. 11 (1981), 95–130, at 126.

83. Collins used this argument in his study of the gravitational radiation controversy. He pointed out that, in accepting the electrostatic calibration measuring technique, Joseph Weber restricted the interpretative flexibility of gravitational radiation results, and chose not to argue on certain fronts which, in principle, were not entirely implausible: Collins, *Changing Order*, op. cit. note 1, 104–06. For Everett Hughes's methodological principle, see S. Leigh Star, 'Introduction: The Sociology of Science and Technology', *Social Problems*, Vol. 35 (1988), 197–205, at 198.

84. For a longer account, see Olazaran (1991), op. cit. note 9, Chapters 4–5, and Olazaran (1993), op. cit. note 9, 386–417.

85. Fleck (1987), op. cit. note 12.

86. Basically, expert systems are composed of a knowledge-base (where knowledge relevant for a certain domain is represented) and techniques for making inferences from that base in a particular situation or problem. In these knowledge-based information processing systems, emphasis is laid on symbolic representation and on the ability of the computer to carry out structure-sensitive transformations of those representations.

87. Fleck (1987), op. cit. note 12, 153.

88. Geoffrey E. Hinton and J.A. Anderson, *Parallel Models of Associative Memory* (Hillsdale, NJ: Laurence Erlbaum, 1981).

89. Dana H. Ballard, G.E. Hinton and Terrence Sejnowski, 'Parallel Visual Computation', *Nature*, Vol. 306 (3 November 1983), 21–26.

90. Jerome A. Feldman and D.H. Ballard, 'Connectionist Models and their Properties', *Cognitive Science*, Vol. 6 (1982), 205–54.

91. Parallel computing entails the use of more than one processor working concurrently on a problem. According to their '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).

92. Researchers have also started to design and use special-purpose hardware for implementing neural nets. For Carver Mead's pioneer work on VLSI analog circuits, see C. Mead, *Analog VLSI and Neural Systems* (Reading, MA: Addison-Wesley, 1989).

93. The two PDP volumes were the manifesto of the new neural-net movement: see David E. Rumelhart, James L. McClelland and The PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1, *Foundations* (Cambridge, MA: MIT Press, 1986), and McClelland, Rumelhart and PDP RG, ibid., Vol 2, *Psychological and Biological Models* (Cambridge, MA: MIT Press, 1986). The PDP Group made an important 'marketing' effort aimed at bringing neural nets back to the AI and cognitive science fields.

94. See, for example: Michael J. Mulkay, 'Three Models of Scientific Development', *Sociological Review*, Vol. 23 (1975), 509–26; Mulkay, G. Nigel Gilbert and S. Woolgar, 'Problem Areas and Research Networks in Science', *Sociology*, Vol. 9 (1985), 187–203.

95. It has been pointed out that Hopfield was a well-recognized physicist who could 'afford' to attempt to make a contribution in an area that had not yet been recognized as a valuable or respectable one: see Anderson & Rosenfeld, op. cit. note 14, 457.

96. J.J. Hopfield, 'Neural Networks and Physical Systems with Emergent Collective Computational Abilities', *Proceedings of the National Academy of Sciences*, Vol. 79 (1982), 2554–58, reprinted in Anderson & Rosenfeld, op. cit. note 14, 460–64. The crucial aspect of Hopfield's contribution, — a consequence of his use of the spin-glass metaphor — was the notion of the 'energy' of a (symmetrically-connected) neural net. 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 at local minima of the energy function. One of the most important properties of this type of net is that it can work as a content-addressable memory so that, under the right circumstances, it will retrieve correct whole patterns when presented with degraded versions of input patterns.

97. Hopfield, op. cit. note 96 (reprinted version), 460.

98. David H. Ackley, G.E. Hinton and T.J. Sejnowski, 'A Learning Algorithm

for Boltzmann Machines', *Cognitive Science*, Vol. 9 (1985), 147–69, reprinted in Anderson & Rosenfeld, op. cit. note 14, 638–49.

99. Sejnowski, interview, 8 November 1989. See also Ackley, Hinton & Sejnowski, op. cit. note 98 (reprinted version), 641.

100. T.J. Sejnowski and Halbert White, Introduction to reprinted version of Nils J. Nilsson, *The Mathematical Foundations of Learning Machines* (San Mateo, CA: Morgan Kaufmann, 1991; original version 1965), vii–xxi, quote at xii.

101. 'The Boltzmann machine is a generalization of the Perceptron to more than one layer. It's interesting, it turned out that the key assumptions you had to change were two things. First that there are feedback connections à la Hopfield, so you have symmetric connections, so it's no longer feedforward net but symmetric net with feedback connections. And second of all, the Perceptron was a deterministic machine, whereas the Boltzmann machine was probabilistic. So you make those two changes, and then suddenly it's a completely different architecture, suddenly you can prove theorems, you can discover learning algorithms, you can solve problems that the Perceptron couldn't', Sejnowski, interview, 8 November 1989.

102. D.E. Rumelhart, G.E. Hinton and Ronald J. Williams, 'Learning Internal Representations by Error Propagation', in Rumelhart, McClelland & PDP RG, op. cit. note 93, 318–62.

103. A learning cycle in a back-propagation net 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. By adjusting the connections of the system according to this technique, the total error measure for a set of input/output patterns is minimized in a gradient descent way. See D.J. Rumelhart, G.E. Hinton and R.J. Williams, 'Learning Representations by Back-propagating Errors', *Nature*, Vol. 323 (9 October 1986), 533–36, reprinted in Anderson & Rosenfeld (eds), op. cit. note 14, 696–99, at 697.

104. Sejnowski & White, op. cit. note 100, xv.

105. Outside the context of the neural-net revival, Paul Werbos's algorithm for multilayer nets was not considered practical by Minsky in the 1970s: see Olazaran (1993), op. cit. note 9, 396–406. For Werbos's contributions see: P.J. Werbos, *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences* (Cambridge, MA: unpublished PhD dissertation, Harvard University, 1974); Werbos, 'Applications of Advances in Nonlinear Sensitivity Analysis', in R.F. Drenick and F. Kozin (eds), *Systems Modelling and Optimization: Proceedings of the 10th IFIP Conference, New York City, 31 August–4 September 1981* (New York: Springer-Verlag, 1982), 762–70.

106. See Richard Forsyth, 'The Brain Mimics Are Back in Business', *The Guardian* (London, 12 January 1989), 25.

107. Rumelhart, Hinton & Williams, op. cit. note 93, 321, 324 & 361 (emphasis in original).

108. The role of the PDP Group in rewriting the official history of the controversy helps explain Grossberg's priority complaints. Grossberg claimed that some of his models were rediscovered in the neural-net revival, and that he did not

receive enough recognition for his previous work: see S. Grossberg, 'Competitive Learning: From Interactive Activation to Adaptive Resonance', *Cognitive Science*, Vol. 11 (1987), 23–63.

109. M.L. Minsky and S.A. Papert, *Perceptrons: An Introduction to Computational Geometry* (Cambridge, MA: MIT Press, 1988), 260–61; this is a second, enlarged edition of Minsky & Papert, op. cit. note 13.

110. See, for example, Olazaran (1991), op. cit. note 9, 274–75.

111. Ibid., 282–92.

112. Ibid., 276–81.

113. Ibid., 293–307.

114. I have adopted these categories from Pinch's case study: see Pinch, op. cit. note 7.

Mikel Olazaran carried out his doctoral research in the Department of Sociology at the University of Edinburgh (UK), receiving his PhD in 1991. He is now a lecturer in sociology at the University of the Basque Country, and a member of the ILCLI research institute at that university. He is conducting research on the sociology and history of AI (he is preparing a book based on his PhD thesis), on the social dimensions of information technology, and on the science and technology system of the Basque-Navarre region.

Author's address: Departamento de Sociologia, Facultad de Ciencias Sociales y de la Comunicación, Universidad del País Vasco, Apartado 644, 48080 Bilbao, Spain. Fax: +34 4 464 8299; e-mail: cipolrom@lg.ehu.es