Computer Modeling in Philosophy

Istvan S. N. Berkeley*

# The Curious Case of Connectionism

https://doi.org/10.1515/opphil-2019-0018
Received April 29, 2019; accepted June 28, 2019

**Abstract:** Connectionist research first emerged in the 1940s. The first phase of connectionism attracted a certain amount of media attention, but scant philosophical interest. The phase came to an abrupt halt, due to the efforts of Minsky and Papert (1969), when they argued for the intrinsic limitations of the approach. In the mid-1980s connectionism saw a resurgence. This marked the beginning of the second phase of connectionist research. This phase did attract considerable philosophical attention. It was of philosophical interest, as it offered a way of counteracting the conceptual ties to the philosophical traditions of atomism, rationalism, logic, nativism, rule realism and a concern with the role symbols play in human cognitive functioning, which was prevalent as a consequence of artificial intelligence research. The surge in philosophical interest waned, possibly in part due to the efforts of some traditionalists and the so-called black box problem. Most recently, what may be thought of as a third phase of connectionist research, based on so-called deep learning methods, is beginning to show some signs of again exciting philosophical interest.

**Keywords:** Connectionism, Neural Networks, History, Philosophy

## 1 Introduction

When it comes to the relationship between computer models and philosophy, the class of models which broadly fall under the heading of ‘connectionism’ have arguably been amongst the most philosophically significant, or perplexing.[1] What makes the case of connectionism ‘curious’ is the fact that it is one of the relatively rare case where innovations within sub-disciplines of cognitive science have had a significant influence in philosophy and generated a good deal of philosophical commentary. The undertaking here is to look at some of the philosophical claims and issues that connectionist research has given rise to. In doing this, the history of connectionism will be broken down into three broad phases, which will be treated chronologically. The first phase had little philosophical impact, but served to set the scene for the second phase, which had more significant philosophical effects. For this reason, it will only be briefly sketched here. Boden provides a much more detailed account.[2] The main focus will instead be on the second phase of connectionism. The paper will conclude with a brief discussion of the third phase, which is beginning to show some philosophical promise. Walker provides a more detailed comparison between the first and second phases of connectionism, and the potential psychological implications.[3]

1 For overviews, please see Dreyfus and Dreyfus, “Making a Mind verses Modeling a Brain: Artificial Intelligence Back at a Branchpoint” and Aizawa, “Connectionism and Artificial Intelligence: History and Philosophical Interpretations”. It is also worth mentioning here that the class of systems here referred to as being ‘connectionist’ are also known as ‘artificial neural network’ systems and even ‘neural network’ systems. The term ‘connectionist’ is used here for terminological consistency.
2 Boden, *Mind as Machine: A History of Cognitive Science*.
3 Walker, “A Brief History of Connectionism and its Psychological Implications”.

***Corresponding author: Istvan S. N. Berkeley,** The University of Louisiana at Lafayette, Lafayette, United States of America; E-mail: Istvan@louisiana.edu

## 1.1 In the beginning…

Broadly speaking, connectionist research focuses on computational systems that are made up of relatively simple processing elements that are interconnected in a manner which enables them to solve tasks which are of interest to the researchers. This is a very general overview, as there are many different connectionist architectures and components. Connectionist architectures are fundamentally different from those found in other types of systems studied by artificial intelligence researchers, such a say production systems.[4] The difference lies in the fact that connectionist system are not explicitly built out of formal structures in quite the same way. Another feature that is commonly associated with connectionism is that it is often claimed, to varying degrees (and with varying amounts of plausibility), that the inspiration for this architectural type derives from the study of biological neural systems.

The research tradition which falls broadly under the term 'connectionism' can be traced back to a publication of a paper in 1943 by Warren McCulloch and Walter Pitts[5], although the term 'connectionism' originated in the work of Thorndyke in the 1930s.[6] McCulloch and Pitts attempted to mimic the function of biological neurons, albeit in a highly simplified manner, using a mathematical formalism. Their artificial neuron had two states, either active, or inactive, depending upon the inputs it received. As a matter of fact, there had been a research tradition roughly along these lines in various fields prior to the McCulloch and Pitts publication. These are discussed by Boden, in a little more detail.[7]

There is some evidence that both McCulloch and Pitts had philosophical interests. Schlatter and Aizawa conjecture that Pitts went to the university of Chicago some time in 1938 in order to meet Bertrand Russell, who was a visiting professor there.[8] They also suggest that Russell put Pitts in touch with Rudolph Carnap. Abraham discusses McCulloch's various philosophical interests and aspirations.[9] McCulloch's book *Embodiments of Mind* also shows considerable philosophical inclinations. Aizawa provides further details.[10] However, despite these philosophical involvements, McCulloch and Pitts' work had a limited philosophical impact, although Piccinini cites their 1943 work as the first use of computation to address the mind-body problem.[11]

Before moving on from the discussion of the work of McCulloch and Pitts, it is also probably worth briefly mentioning that in 1947 Pitts and McCulloch published a paper with the title, "How We Know Universals the Perception of Auditory and Visual Forms." The language of the title alone suggests a highly philosophical orientation. However, as Aizawa notes, this paper is somewhat puzzling, as they offer no explicit discussion of what they consider universals to be. Aizawa summarizes their implicit view of universals to amount to something like "…to know a universal is to be able to respond to that universal, while ignoring irrelevant features of particular instances of that universal."[12] It is certainly the case, though, that this paper also had very little impact on philosophical discourse.

## 1.2 Perceptrons

Whereas the work of McCulloch and Pitts was firmly rooted in attempts to understand aspects of neural circuitry, in a formal and mathematical way, the work of the next important figure, Frank Rosenblatt, was much more centrally rooted in psychology.

---

**4** For the technical details concerning production systems, see Klahr, Langly and Neches, *Production System Models of Learning and Development*.
**5** McCulloch and Pitts, "A logical calculus of the ideas immanent in nervous activity".
**6** Thorndyke, *The Fundamentals of Learning*.
**7** Boden, *Mind as Machine: A History of Cognitive Science*, 1121–1123.
**8** Schlatter and Aizawa, "Walter Pitts and 'A Logical Calculus'", 238.
**9** Abraham, Rebel Genius: *Warren S. McCulloch's Transdsciplinary Life in Science*.
**10** Aizawa, "Connectionism and Artificial Intelligence: History and Philosophical Interpretations".
**11** Piccinini, "The First Computational Theory of Mind and Brain: A Close Look at McCulloch and Pitts 'Logical Calculus of Ideas Immanent in Nervous Activity'".
**12** Aizawa, "Warren McCulloch's Turn to Cybernetics: What Walter Pitts Contributed", 213.

While working at the Cornell Aeronautical Laboratory, in 1958 Frank Rosenblatt published “The Perceptron: A Probabilistic Model For Information Storage and Organization in the Brain” in *Psychological Review*. In this paper Rosenblatt proposes to address questions of how information is stored and remembered and how this information influences recognition and behavior in ‘higher organisms’. This paper is of interest because Rosenblatt explicitly describes the answers to the questions he intends to explore as being ‘connectionist’.[13] He is also far from shy about emphasizing the potential of perceptrons. In the concluding sentence of the paper, Rosenblatt states,

> By the study of systems such as the perceptron, it is hoped that those fundamental laws of organization which are common to all information handling systems, machines and men included, may eventually be understood.[14]

Rosenblatt followed McCulloch and Pitts in using processors which were all-or-none, in terms of their activation state. A focus on systems built up of components of this kind is one of the hallmarks of what is identified here as the first phase of connectionist research. However, Rosenblatt’s main innovation beyond the work of McCulloch and Pitts was to design various systems that could undergo training. Later, in 1962, Rosenblatt extended this research in his monograph *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*.

Rosenblatt’s Perceptrons consisted of banks of processing units arranged in a couple of layer, one of which acted as inputs to the system, the other acting as the outputs of the system, with modifiable weighted connections between the various components. Training took place by changes being made to these weighted connections. Of particular significance was the fact that Rosenblatt was able to prove the so-called ‘Perceptron Convergence Theorem’ for his systems. In layman’s terms, this theorem showed that, if there was a solution to a particular problem, then within a finite number of steps, a perceptron would be able to find a solution to the problem, such that it gave correct responses for all the training patterns.[15]

## 1.3 *Perceptrons*

Rosenblatt’s work attracted quite a bit of media attention, which had both advantages and disadvantages, according to Boden.[16] In particular, Hecht-Nielsen (quoted in Anderson and Rosenfield) noted that “[t]he wall-to-wall media coverage of Rosenblatt and his machine irked Minsky.”[17] A similar claim was made by Feldman in a talk at Buffalo, given on 26th July 1994. This is of significance, as the next phase of this story has Minsky and his colleague Papert front and center.

Minsky was one of the early pioneers of artificial intelligence research. He, along with John McCarthy, co-founded what is now known as the Computer Science and Artificial Intelligence Laboratory, at MIT.[18] Although Minsky initially studied connectionist style systems, he eventually focused upon exploring intelligence as based upon symbolic manipulation. This approach is often characterized as being a contrast class to connectionism.[19]

As Hecht-Nielson recounts the events surrounding the publication of Minsky and Papert’s famous 1969 book *Perceptrons,* it was the culmination of a concerted campaign against Rosenblatt to ensure funding for their Artificial Intelligence research.[20] Hecht-Nielson even compares Minsky to Darth Vader![21] The main point of *Perceptrons* was to draw attention to the limitations of perceptrons. They argued that, assuming some not

13 Rosenblatt, “The Perceptron: A Probabilistic Model For Information Storage and Organization in the Brain”, 387.
14 Ibid., 407–408.
15 See Rosenblatt, *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*.
16 See Boden, *Mind as Machine: A History of Cognitive Science*, 918-919.
17 Anderson and Rosenfield, *Talking Nets: An Oral History of Neural Networks*, 304.
18 See https://www.csail.mit.edu/
19 See for example Fodor and Pylyshyn, “Connectionism and Cognitive Architecture: A Critical Analysis”.
20 See Anderson and Rosenfield, *Talking Nets: An Oral History of Neural Networks*, 304–307.
21 Ibid., 304–305.

unreasonable assumptions, such as limiting the conclusions to networks without massive interconnectivity between processing units, perceptrons could not handle connectedness, counting objects, or even learning the truth-table for XOR.

In many ways, the Hecht-Nielson account of the publication of *Perceptrons* paints Minsky and Papert as pantomime villains, straight out of central casting. A more nuanced and detailed account of events appears in Boden.[22] Medler characterized the situation as follows: "[t]he publication of *Perceptrons* in 1969 by Minsky and Papert has taken on a mythical aura – it has been likened to the huntsman being sent out to bring back the heart of Snow White."[23]

The claims about funding for Artificial Intelligence (AI) appear to be false. According to Minsky, their research was "...very well funded, we had $1,000,000 per year from ARPA from 1963 to 1970, and more later. You can't prove theorems for political purposes, anyway."[24] In an admirable turn of phrase, Minsky continues, "It would seem that *Perceptrons* has much the same role as *The Necronomicon* – that is, often cited but never read."[25] Nonetheless, the publication of *Perceptrons* had a significantly chilling effect upon connectionist research. Most notably, according to Boden, it prevented funding of this line of research from major funding agencies.[26] This effectively put the bulk of connectionist research (with a few notable exceptions) into a state of hibernation until the mid-1980s when the second phase of connectionism took off. It was this second phase of connectionism that held the greatest philosophical promise. The decline in interest in connectionism marked the end of the first phase of connectionist research.

## 2 Parallel Distributed Processing (PDP)

The key event which kicked off the second phase of connectionist research was the publication of a two volume work *Parallel Distributed Processing: Explorations in the Miccrostructure of Cognition*. The importance of these two volumes can be estimated by the fact that Davis refers to this work simply as the PDP 'Bible'.[27] This work was attributed to Rumelhart, McClelland and The PDP Research Group. The members of this group are listed by Rumelhart, McClelland *et al.*.[28] Although this list contains people from various disciplines, such as psychology, computer science, cognitive science, and even linguistics, it is interesting to note in the current context that no philosophers are included in this list. Boden notes that the Rumelhart, McClelland, *et al.* text was designed to be accessible, in order to attract interest, especially from graduate students. Steps were also taken to bring down the cost to maximize accessibility. Boden also notes that, "[t]hey hoped to have well-thumbed, coffee-stained, volumes on student desks–not pristine pages sitting on library shelves."[29]

The systems discussed in this volume were fundamentally parallel, as they were composed of many comparatively simple processors. They were also distributed as, rather than there being a single processor, processing took place across multiple processing elements that worked together to determine the final output of the systems. In a sense, these systems were similar to perceptrons in these respects. However, the significant innovation of these systems was that, rather than just having input and output units, the systems they described also had intermediate processors, the so-called 'hidden layers', which were not directly connected to the inputs, or outputs.

The main technical innovation which launched the second phase of connectionist research was the development of processing elements which had activation functions which were more complex than

**22** Boden, *Mind as Machine: A History of Cognitive Science*, 911–916.
**23** Medler, "A Brief History of Connectionism", 77.
**24** Minsky, personal communication, 15th of August, 1994.
**25** *The Necromonicon* is a fictional grimoire, or book of magic that appears in the works of H. P. Lovecraft. The putative author is Abdul Alhazred. See Lovecraft "H. P. Lovecraft, The Fiction, Complete and Unabridged".
**26** Boden, *Mind as Machine: A History of Cognitive Science*, 914.
**27** Davis, "Two Notions of Implicit Rules", 153.
**28** Rumelhart, McClelland et al. *Parallel Distributed Processing: Explorations in the Miccrostructure of Cognition*, xix–xx.
**29** Boden, *Mind as Machine: A History of Cognitive Science*, 947.

the simple all-or-none processors of the Rosenblatt era. This in turn facilitated the development of more complex systems of processing units, with more layers of processors, namely the hidden units. A commonly used processor type had a sigmoid activation function. The sigmoid activation function had roughly an 'S' shape. So-called Radial Basis Function (RBF) units, which had a Gaussian activation were also introduced. The additional mathematical complexity of these activation functions made it possible to take derivatives, which in turn facilitated training these more complex networks. To a large degree, much of the work in Rumelhart and McClelland *et al*. volume built upon the work of Hopfield, who initially explored the mathematical properties of sigmoid activation functions.[30]

Another important milestone in the reemergence of connectionist research, which is often overlooked in standard accounts, was the publication in 1988 of McClelland and Rumelhart's, *Explorations in Parallel Distributed Processing*. The unique thing about this book was that it came with two floppy disks, which contained versions of many of the systems discussed in the 1986 Rumelhart, McClelland *et al*. book. This was a masterful stroke, as it made connectionist software available to anybody who had access to a computer. While it is one thing to read about the abilities of classes of systems in the abstract, it is something completely different to be able to play with the systems and let people 'get their hands dirty' with the software.

To appreciate why the Rumelhart, McClelland *et al*. publication from 1986 had such a significant impact, it is worth briefly sketching the philosophical context in which it arose. Authors such as Dreyfus and Dreyfus, Aizawa and Boden, to name a few, associated traditional AI research with certain philosophical inclinations.[31] These include conceptual ties to the philosophical traditions of atomism, rationalism, logic, nativism, rule realism and a concern with the role symbols play in human cognitive functioning. This conceptual viewpoint is nicely typified in Haugeland's 1985 book *Artificial Intelligence: The Very Idea*. In this book, Haugeland christens this conceptual view point as 'GOFAI', standing for 'Good Old Fashioned Artificial Intelligence'. The problem was that at the time Haugeland's book came out GOFAI was in trouble. Early promising successes appeared to lead nowhere in particular. Thus, when Rumelhart, McClelland *et al*. published their two volume set, people were looking around for alternatives to GOFAI. In addition, philosophers who were opposed to the philosophical inclinations of GOFAI, were naturally drawn to potential alternatives. These then are likely the reasons that the second phase of connectionism began to have impacts in philosophy.

As a matter of fact, connectionist research had attracted philosophical attention, even prior to the publication of the PDP volumes. For instance, in 1985 Bechtel published an early philosophical paper on connectionism.[32] A somewhat crude, but nonetheless quite effective means of gauging the impact of connectionism on philosophy can be gained by querying the *Philosopher's Index* and counting the number of hits it gives, by year. This was done looking for the terms 'connectionist' and 'connectionism' and the number of entries found were then tabulated and recorded. The results are recorded in Figure 1. The specific version of the *Philosophers' Index* used was the on-line version available through EBSCOHOST. For the sake of completeness, it is worth mentioning that the terms 'Parallel Distributed Processing' and 'PDP' were also checked, but produced too few results to be worth reporting.

A brief inspection of Figure 1 makes it abundantly clear that the period of peak interest in connectionism in the philosophical literature was the mid to late 1990s. The raw data upon which this plot was based is presented in Table 1.

While these results have heuristic value, they should not be construed as being canonical, for a number of reasons. First, during this period of time a number of new journals dealing mainly with connectionist themes, such as *Connection Science* and *Neural Computation* were founded. Both of these were founded in 1989. However, no papers in *Neural Computation* and just one paper in *Connection Science* are included in

**30** For further details, see Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational Abilities" and Hopfield, "Neurons with Graded Response have Collective Computational Properties LikeThose of Two-State Neurons".

**31** See Dreyfus and Dreyfus, "Making a Mind verses Modeling a Brain: Artificial Intelligence Back at a Branchpoint", Aizawa, "Connectionism and Artificial Intelligence: History and Philosophical Interpretations" and Boden, *Mind as Machine: A History of Cognitive Science*.

**32** Bechtel, "Contemporary Connectionism: Are the New Parallel Distributed Processing Models Cognitive or Associationist?"

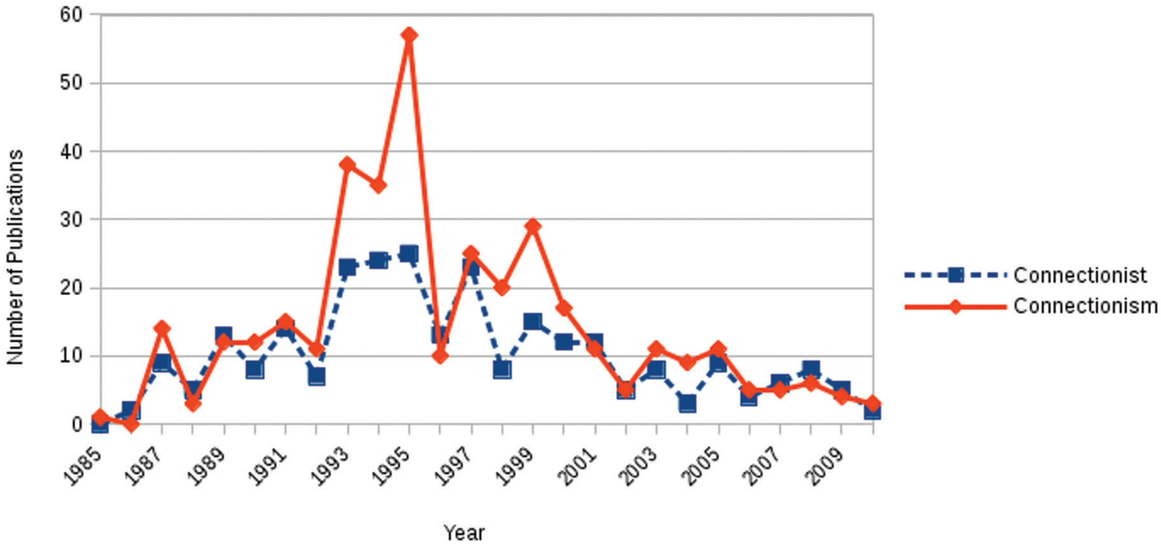


**Figure 1:** The number of instances, by year, that the terms 'Connectionist' and 'Connectionism' appear in searches of *The Philosopher's Index*, for the twenty-five year period from 1985 to 2010.

**Table 1:** The number of occurrences of the terms 'Connectionist' and 'Connectionism' in listings in *The Philosopher's Index* by year, over the twenty-five years between 1985 and 2010.

| Term | 'Connectionist' | 'Connectionism' |
|---|---|---|
| **Year** | | |
| 1985 | 0 | 1 |
| 1986 | 2 | 0 |
| 1987 | 9 | 14 |
| 1988 | 5 | 3 |
| 1989 | 13 | 12 |
| 1990 | 8 | 12 |
| 1991 | 14 | 15 |
| 1992 | 7 | 11 |
| 1993 | 23 | 38 |
| 1994 | 24 | 35 |
| 1995 | 25 | 57 |
| 1996 | 13 | 10 |
| 1997 | 23 | 25 |
| 1998 | 8 | 20 |
| 1999 | 15 | 29 |
| 2000 | 12 | 17 |
| 2001 | 12 | 11 |
| 2002 | 5 | 5 |
| 2003 | 8 | 11 |
| 2004 | 3 | 9 |
| 2005 | 9 | 11 |
| 2006 | 4 | 5 |
| 2007 | 6 | 5 |
| 2008 | 8 | 6 |
| 2009 | 5 | 4 |
| 2010 | 2 | 3 |

the *Philosopher's Index*.[33] So, some philosophical work concerning connectionism will be overlooked.[34] A second reason that the *Philosopher's Index* cannot be taken as a canonical source derives from the fact that, given the interdisciplinary nature of connectionism, important publications appear in volumes which are not in the purview of the *Philosopher's Index*. Ramsey, Stitch and Rumelhart is a good example of this.[35] Indeed, probably the most philosophically significant publication concerning connectionism, Fodor and Pylyshyn, originally appeared in the journal *Cognition*.[36] It only makes an entry into the *Philosopher's Index*, when it was reprinted in McDonald and McDonald.[37] Irrespective of the details though, what is clear is that there was an explosion of philosophical interest in connectionism following the publication of the Rumelhard, McClelland *et al.* volumes.

## 2.1 Fodor and Pylyshyn (1988)

Although philosophical interest was piqued by the arrival of the second generation of connectionism, not all the interest was positive. Somewhat notoriously, Fodor and Pylyshyn published a long paper in *Cognition* arguing that connectionist models would fail to be able to handle systematicity and compositionality, unless they amounted to implementational variants of traditional 'classical' AI type systems. Systematicity is the property of a representational system which supposedly ensures that there are certain symmetries in the representational system such that the ability to represent something will ensure the ability to also represent other things with related semantic content. Compositionality is the idea that the meaning of a complex expression is a function of the constituents parts of the expression, along with the rules for combining those parts. So, to borrow one of Fodor and Pylyshyn's examples, if a system can handle the sentence 'John loves Mary', it should also be able to handle the sentence 'Mary loves John'. Their claim is that classical systems can do this with ease, while connectionist ones cannot, unless they are implementational variants of classical systems.

Rhetorically, this was an inspired move. In 1984, Pylyshyn had argued that cognitive systems could be usefully discussed using three distinct levels of description.[38] In this, Pylyshyn broadly followed Marr's three levels of analysis of cognitive systems.[39] Marr's three levels were the computational level, the algorithmic level and the implementational level. However, it was axiomatic to Pylyshyn's position that the implementational level of description was non-cognitive. So, Fodor and Pylyshyn's conclusion, if accepted, would preclude connectionist systems from playing any fruitful, or useful, role in cognitive science.

For Fodor, connectionism represented a potential threat to his research program. Famously, Fodor had argued that there had to be a 'language of thought', with certain particular properties.[40] His argument for this view was that this was 'the only game in town' (paraphrasing Richard Nixon). The properties that Fodor attributed to the language of thought were very similar to the kinds of properties that were found in both natural languages and many traditional approaches in Artificial Intelligence research. However, the rise of connectionism, at least potentially, suggested that there may be another 'game' possible. So, Fodor and Pylyshyn's publication represented a defense of the traditional philosophical view that derived from the AI research tradition.[41] So, the publication from Fodor and Pylyshyn (1988) was one of the early salvos in a philosophical debate between connectionists and more traditional views.

**33** This single paper is Waskan, "A Critique of Connectionist Semantics".
**34** See for example Clark "A review of "Simple Minds" by D. Lloyd, 1989, MIT Press, London", Clark "Representation, Development and Situated Connectionism", Chalmers "Connectionism and Compositionality: Why Fodor and Pylyshyn were wrong", and Berkeley, Dawson et al. "Density Plots of Hidden Unit Activations Reveal Interpretable Bands".
**35** Ramsey, Stitch and Rumelhart, *Philosophy and Connectionist Theory*.
**36** Fodor and Pylyshyn, "Connectionism and Cognitive Architecture: A Critical Analysis".
**37** McDonald and McDonald, *Connectionism: Debates in Psychological Explanation*.
**38** Pylyshyn, *Computation and Cognition: Towards a Foundation for Cognitive Science*.
**39** Marr, Vision: *A Computational Approach*.
**40** Fodor, *The Language of Thought*.
**41** C.f. Haugeland, *Artificial Intelligence: The Very Idea*.

Another reason that Fodor and Pylyshyn's paper is important is that it raised questions and topics which served to stimulate a good amount of both philosophical and technical commentary and debate. Of particular note was the lively philosophical debate that arose between Aizawa and Hadley concerning systematicity, productivity, and cognitive architecture. Aizawa provides a helpful summary of the various moves in this debate.[42] Another philosophical response can be found in Van Gelder. He endorses a position that is neither unequivocally connectionist, nor traditional. Rather, he attempts to pursue something like a middle path.[43] However, it was clear that the rise of connectionism succeeded in opening a proverbial philosophical 'can of worms'.

There was also push-back against Fodor and Pylyshyn, with more technical methods. Of particular note were the responses from Smolensky and Pollack.[44] They both offered formal approaches that seemed to undermine the core claims advanced by Fodor and Pylyshyn. However, such challenges were not well received by Fodor, at least. Others, such as Chalmers (1993) and Dawson, Medler and Berkeley (1997) combined both philosophical and technical aspects.[45] So, despite all the drama, the second generation of connectionist systems opened up the philosophical arena to a wider range of philosophical positions than had seemed plausible prior to these innovations.

Fodor warmed to his crusade against connectionism. In 1990, he followed his paper with Pylyshyn, with a paper with McLaughlin (1990).[46] In this paper, Fodor and McLaughlan took specific issue with the proposals made by Smolensky in 1987, arguing that one proposal was too weak to be relevant and that the other failed to properly account for systematicity.[47] Even after this, Fodor was not finished. In 1997 he published another paper with the provocative title, "Discussion: Connectionism and the Problem of Systematicity (continued): Why Smolensky's Solution Still does not Work." The philosophical debate that the original Fodor and Pylyshyn paper helped to generate explains, at least in part, how philosophers began to be interested in connectionist research.

## 2.2 The rise of connectionism in philosophy

As was noted earlier, connectionism provided an interesting alternative to the atomistic, rule based ideas that were prevalent in philosophy at the time. In 1989, two philosophical books were published which also helped to bring connectionism into the philosophical mainstream. The first of these was Lloyd's (1989) book, *Simple Minds*. This work at least took the potential contribution of connectionism to the philosophy of mind seriously. However, probably of greater significance was the publication in 1989 of Clark's *Microcognition*. The big advantage of Clark's book was that it offered a discussion of connectionism which was reasonably detailed, but did not get too bogged down with technicalities. This book was thus highly accessible to interested philosophers. Clark also made the inspired choice of introducing connectionism to philosophers mostly through a discussion of the Jets and Sharks model, originally published by McClelland, Rumelhart and Hinton.[48] The Jets and Sharks model took the properties of the two gangs from the 1961 film *West Side Story* and mapped names onto gang affiliation, age, level of education, marital status and occupation. While the Jets and Sharks model did not learn and was in fact constructed by hand, it had two advantages. First, it was reasonably intuitive and easy to understand, even by novices. Second, an implementation of this model was included with the software supplied with the 1988 McClelland, and Rumelhart text. So,

**42** Aizawa, *The Systematicity Arguments*.
**43** Van Gelder, "Compositionality: A Connectionist Variant on a Classical Theme".
**44** Smolensky, "Tensor Product Variable Binding and the Representation of Symbolic Structures in Connectionist Systems" and Pollack, "Recursive Distributed Representations".
**45** Chalmers, "Connectionism and Compositionality: Why Fodor and Pylyshyn were Wrong" and Dawson, Medler and Berkeley, "PDP Networks Can Provide Models That Are Not Mere Implementations of Classical Theories".
**46** Fodor and McLaughlin, "Connectionism and the Problem of Systematicity: Why Smolensky's Solution Doesn't Work".
**47** Smolensky, "The Constituent Sructure of Mental States: A Reply to Fodor and Pylyshyn".
**48** McClelland, Rumelhart and Hinton, "The Appeal of PDP".

interested philosophers, or even skeptical philosophers, had a reasonably easy means to test the various claims that have been advanced.

In *Microcognition*, Clark was also able to avail himself of the relatively novel conceptual framework that connectionism seemed make appear plausible. He argued for a view of mind that was much less atomistic and decidedly more holistic, as compared to many of the alternatives. Other early advocates of the importance of connectionism to the philosophy of mind included Horgan and Tienson, Bechtel, Bechtel and Abrahamsen, Paul Churchland, and Patricia Churchland and Sejnowski, to name just a few.[49]

This era also had various meetings where connectionism was a major topic of discussion. For instance, a number of the papers that appear in Horgan and Tiensen's 1991 collection were originally presented at the Spindel Conference held in 1987 at Memphis State University.[50] Similarly, the annual Vancouver Studies in Cognitive Science Conference held at Simon Fraser University in 1990 was focused on the topic of connectionism. Some of the papers from this conference subsequently appeared in edited collection by Stephen Davis.[51] Interestingly, there is also a set of notes taken at this conference by graduate students.[52] Contributions concerning connectionism also began to appear on the schedules of both national and regional philosophy conferences, during this period.

Suffice to say, by the early 1990s, connectionism was getting a great deal of interest, both in philosophy and in other disciplines. Indeed, as a movement, connectionism was taking on an almost revolutionary tone. In 1987, Schneider raised the question whether connectionism amounted to a fundamental paradigm shift in psychology.[53] Also in 1987, Horgen and Tienson echoed this sentiment when they noted, "There is a Kuhnian crisis in GOFAI, brought on by a pattern of unfulfilled promises and disappointing results."[54] In 1991, they were to explicitly argue the case for this Kuhnian interpretation.[55]

Another important, though often overlooked, event which helped propel connectionism was an unusual event which rejoiced in the name FISI-CS. This stood for The First International Summer Institute in Cognitive Science. It was held at SUNY Buffalo from July 5th to 30th, 1994. It was a remarkable event with around 400 participants from 32 countries and about 100 faculty and plenary speakers. Participants were, for the most part, graduate students and faculty, with a few undergrads drawn from universities and industries from all over the world.[56]

The institute offered a range of classes, in addition to a plenary speaker most weekday evenings. The formal presentations were followed by more informal so-called 'fire-side' chats. Over the weekends there were workshops and social events.[57] The reason this is relevant here is because one of the courses was about connectionism. It was taught by Rumelhart and Smolensky. The course description said explicitly, "This course assumes no prior background in connectionism (or linguistics)." In other words, FISI-CS provided a perfect vehicle to proselytize the cause of connectionism. Needless to say, the Rumelhart and Smolensky class was very well attended. In addition, there was a two day workshop on "Connectionism and Neuroscience" and a two day workshop on just connectionism, which featured Smolensky, Hinton, McClelland and Feldman. Although participants came from a range of disciplinary backgrounds, there was

**49** See Horgan and Tienson, *Connectionism and The Philosophy of Mind*, Bechtel "Contemporary Connectionism: Are the New Parallel Distributed Processing Models Cognitive or Associationist?", Bechtel and Abrahamsen, *Connectionism and the Mind: An Introduction to Parallel Processing in Networks*, Paul Churchland, "On the Nature of Theories: A Neurocomputational Perspective", and Patricia Churchland and Sejnowski, "Neural Representation and Neural Computation".

**50** See Horgan and Tienson, *Connectionism and The Philosophy of Mind* and Horgan and Tienson, "Spindel Conference 1987: Connectionism and The Philosophy of Mind".

**51** Davis, *Connectionism: Theory and Practice*.

**52** Available at https://www.downes.ca/post/68976.

**53** Schneider, "Connectionism: Is it a Paradigm Shift for Psychology?".

**54** Horgan and Tienson, "Spindel Conference 1987: Connectionism and The Philosophy of Mind", 2.

**55** Horgan and Tienson, "Settling into a New Paradigm", 241–260.

**56** For an incomplete list of participants, faculty and plenary speakers see http://ftp.cse.buffalo.edu/users/rapaport/fisi-addresses.html.

**57** A detailed schedule of classes, talks and workshops appears in the FISI Handbook, which is available at https://cse.buffalo.edu/~rapaport/Papers/FISI-Handbook.pdf.

a significant contingent of philosophers. Thus, connectionist ideas became easily accessible to anyone with an interest, including philosophers.

At this point it is worth turning attention more directly to looking at the specific philosophical claims which connectionism was (at least putatively) supposed to support. As the literature is pretty huge, rather than attempting to be encyclopedic, the approach that will be adopted here is to focus on particular philosophical claims, which arose with some regularity.

## 2.3 Nativism and innateness

Broadly speaking, nativism, or innateness is a view that important parts of the human epistemological systems, or cognitive architecture, must be naturally supplied to us, by some process (the exact answers differ here), rather than being simply learned from experience. Nativism, in one form or another, has been a philosophical position that has historically been through periods of popularity and decline.[58] Traditionally, nativist positions have been advocated for by philosophers such as Plato, Descartes and Leibniz. As such, the view has became associated with rationalistic views of things. By contrast, more empirically inclined philosophers, such as Locke and Hume have battled against the nativist tradition.

In recent times, nativism has been advocated for in linguistics, by Chomsky (1965).[59] In the philosophy of mind, nativism was famously championed by Fodor (1983).[60] As many of the arguments concerning nativism have been focused upon human linguistic capacities, connectionist simulations that focused upon language were thus a potential source of anti-nativist arguments and evidence.

An important development concerning the relationship between connectionism and innateness occurred in 1996, when Elman, Bates *et al.* published a volume with the title *Rethinking Innateness*.[61] This book amounted to a sustained attack upon many commonly held nativist views. Interestingly enough, this volume was followed in 1997 by Plunkett and Elman's work, which included the T*learn* suite of software, which provided a variety of connectionist simulations, which broadly supported the claims made in the book of the previous year.[62] The text was accompanied by software for both Windows and Macintosh. In some senses, this strategy of offering up a theoretical work, which was then supported by software simulations, mirrored that pioneered by Rumelhart, McClelland *et al.* with McClelland and Rumelhart's *Explorations* text, a decade before. However, from a rhetorical point of view, making claims concerning nativism and innateness, supported by simulations, was more persuasive than merely offering arguments. So, with this, connectionism was again being used to attack the standard GOFAI inspired philosophical perspective. This was not the only front where connectionism acted as a disruptive influence.

## 2.4 The question of rules

One claim which commonly arises in the philosophical literature concerning connectionism is that in some sense networks provide an alternative to the systems of rules which often characterize work in GOFAI. For instance, Bechtel and Abrahamsen, while discussing a logic problem task, remark "…we raise the possibility that the rules can be eliminated entirely in the modeling medium, letting networks do all the work."[63] Quite what the nature of the work the network was supposed to be doing, or how it is supposed to be doing it, is unclear. In fact, issues and questions concerning rules pretty rapidly gets really quite complicated. This is,

58 See Samet, "The Historical Controversies Surrounding Innateness".
59 Chomsky, *Aspects of a Theory of Syntax*.
60 Fodor, *The Modularity of Mind*.
61 Elman, Bates, Johnson, Karmiloff-Smith, Parisi and Plunkett, *Rethinking Innateness: A Connectionist Perspective on Development*.
62 Plunkett and Elman, *Exercises in Rethinking Innateness: A Handbook for Connectionist Simulations*.
63 Bechtel and Abrahamsen, *Connectionism and the Mind: An Introduction to Parallel Processing in Networks*.

in part, because of the large philosophical literature concerning rules and rule following.[64] However, it is clear that this line of thinking is common in connectionist inspired philosophy. To cite a similar example, Clark remarks "The PDP model challenges this assumption that in-the-head mechanisms mirror structured, componential, rule-based linguistic theories."[65] So, connectionism offered the promise of considering cognition from a perspective that was markedly dissimilar to that proposed by GOFAI.

Perhaps the best known of the rule skeptics are Horgan and Tienson.[66] Their view was subject to philosophical doubts, objections and questions.[67] There were also a further technical problems which arose from claims like these: there was simply a lack of any direct evidence. While in a program say running LISP, one could point to the program and say, "Look, there is the rule!" In the case of a connectionist network, there was nothing which could so easily be identified. Should one look at matrices of weights, activations of units, or something else? While there were many potential candidates, the correct one to look at was far from clear. Also, upon what basis could a choice of one potential candidate over another be justified? Some theorists took the fact that there were no obvious rules to count as evidence that there really were no rules.[68] Bechtel and Abrahamsen remark "Without a detailed analysis of the activities of the hidden units (which was not performed), we cannot determine exactly how the network solved this [logic] problem."[69] A likely reason that this analysis was not performed will be discussed below. Before addressing this issue a few words about representations are in order.

## 2.5 Symbols and sub-symbols

One of the foundational tenets of GOFAI is the Physical Symbol System Hypothesis (PSSH), introduced in 1976 by Newell and Simon. This amounted to the hypothesis that a physical symbol system had the necessary and sufficient means to exhibit general intelligent action.[70] Physical Symbol Systems are most naturally understood as the kinds of representations and operations found in traditional GOFAI research. The question of whether connectionist systems met the conditions of being physical symbol systems, in the relevant sense is a vexing one. Newell is clear he thought not, while Simon thought they were physical symbol systems.[71] Berkeley provides a more detailed discussion of these points.[72]

As a general rule, connectionist theorists prefer to consider connectionist systems as operating at the sub-symbolic level, following the terminology introduced by Smolensky. As Smolensky uses the term, sub-symbols are supposed to be the fine grained dynamical features that are below the conceptual level. He conceived the sub-symbolic domain as being between that of the symbolic level and the neural level.[73] This, then, was another way in which connectionism represented an alternative to the GOFAI paradigm. Of course, just as the notion of symbolhood was not entirely conceptually clear, particularly across disciplines, so the idea of what actually constituted a sub-symbol was to some degree equally occult.[74] However, this difference, whatever it may ultimately turn out to be, represented another way in which connectionism potentially provided a philosophical alternative to the received orthodoxy. Clark and Lutz put the matter succinctly when they claim that "Connectionist models...differ from those of conventional AI in (amongst

**64** See Miller and Wright, *Rule-Following and Meaning*, for example.
**65** Clark, *Microcognition*, 163.
**66** Horgan and Tienson, "Representations Without Rules".
**67** See for example, Aizawa, "Representations without Rules, Connectionism and the Syntactic Argument".
**68** Such as Bechtel and Abrahamsen, *Connectionism and the Mind: An Introduction to Parallel Processing in Networks*.
**69** Ibid., 171.
**70** Newell and Simon, "Computer Science as Empirical Inquiry", 116.
**71** See Newell, "Physical Symbol Systems", 171 and Vera and Simon, "Reply to Touretzky and Pomerleau: Reconstructing Physical Symbol Systems".
**72** Berkeley, "What the <0.70, 1.17, 0.99, 1.07> is a Symbol?"
**73** Smolensky, "On the Proper Treatment of Connectionism".
**74** See Berkeley, "What the #$*%! Is a Subsymbol?" for a further discussion.

other things) appearing to operate without traditional symbolic data structures over which computational operations may be defined."[75]

There was, though, a significant problem which arose concerning many of the philosophical claims put forward on the basis of connectionist research. So, it is now time to address this issue.

## 2.6 The Black Box Problem

The Black Box Problem with connectionist systems is that once they have learned to perform a task with an acceptable level of proficiency, it is far from straightforward to determine how they are solving the task at hand. This is not a new problem. For instance, Mozer and Smolensky remark that, "...one thing that connectionist networks have in common with brains is that if you open them up and peer inside, all you can see is a big pile of goo."[76] A similar, though more overtly pessimistic point is made by McCloskey when he remarks that "...connectionist networks should not be viewed as theories of human cognitive functions, or as simulations of theories, or even as demonstration of specific theoretical points."[77]

The point here is that if there is no detailed understanding of how trained networks solve problems, then they have very little that they can contribute to cognitive theorizing, or for that matter to philosophical theorizing. This, then, is the black box problem and it is a serious one.

Back in the early 1990s, there was some skepticism about solving the black box problem. Hecht-Nielson suggested it would require a profound revolution in information processing.[78] Robinson went further suggesting that, "We may have to accept the inexplicable nature of mature networks."[79] Fortunately, such skepticism was unfounded.

A number of techniques have been proposed for analyzing trained networks. For example, a range of approaches are presented by Browne.[80] Unfortunately, many of the methods proposed were subject to frustrating limitations. For instance, Bullinaria showed how the commonly deployed method of cluster analysis can mis-classify exceptional items and miss other crucial aspects of network performance. He also addresses the limitations that can arise from another commonly proposed technique, principle component analysis. Under certain conditions, Bullinaria argues that principle component analysis may fail to capture enough variance to recognize anything but the grossest features.[81] Similarly, the Berkeley, Dawson *et al.* (1995) banding analysis[82] method only apparently worked with relatively obscure so-called 'value units'.[83] Eventually, after some mathematical insights into the relationship between value units and more traditional sigmoid processing units, the banding analysis method was generalized to more traditional units, by Berkeley and Gunay.[84] Berkeley and Raine were even able to show how banding analysis could be used to generate an entirely novel theory of how to solve a particular problem.[85] However, these innovations had little philosophical impact, as philosophical interest in connectionism had passed, even if some interest remained in psychology and cognitive science.[86]

The precise reasons for the sudden rapid rise in philosophical interest in connectionism, followed by a rapid decline is unclear. Perhaps it may be a topic of interest for future historians of ideas. However, the

**75** Clark and Lutz, *Connectionism in Context*, 12.
**76** Mozer and Smolensky, "Using Relevance to Reduce Network Size Automatically", 3.
**77** McCloskey, "Networks and Theories: The Place of Connectionism in Cognitive Science", 387.
**78** Hecht-Nielson, *Neurocomputation*, 10.
**79** Robinson, "Implications of Neural Networks for How We Think about Brain Function", 655.
**80** Browne, *Neural Network Analysis, Architectures and Applications*.
**81** Bullinaria, "Analyzing the Internal Representations of Trained Neural Networks".
**82** Berkeley, Dawson, Medler Schopflocher and Hornsby, "Density Plots of Hidden Unit Activations Reveal Interpretable Bands".
**83** For the technical details of value units see Dawson and Schopflocher "Modifying the Generalized Delta Rule to Train Networks of Non-Monotonic Processors for Pattern Classification".
**84** Berkeley and Gunay, "Conducting Banding Analysis with Trained Networks of Sigmoid Units."
**85** Berkeley and Rayne, "An Old Fashioned Connectionist Approach to a Cajun Chord Change Problem".
**86** See Dawson, *Minds and Machines: Connectionism and Psychological Modeling*, for instance.

second generation of connectionist research did serve to open up the range of philosophical positions that could plausibly be defended, based upon simulation results.

Before attempting to draw some final conclusions about the impact of connectionism on philosophy, for the sake of completeness, it is worth briefly looking at what is happening with connectionist research these days, in what might be thought of as the third phase of connectionism, the era of so-called ‘deep learning’.

# 3 Deep learning

Although the frenzy of interest in connectionism in philosophical circles has waned, it has not stopped completely. Philosophical papers on connectionism still appear from time to time.[87] At the same time, technical research into connectionist systems has continued. Unlike the sharp transition between the first and second phases of connectionist research, the move to the third phase has been more gradual and incremental. The modern incarnation of connectionism takes the form of so-called ‘Deep Learning’.[88]

Deep learning is the third phase of connectionism, yet has garnered limited philosophical interest, so far. However, there are beginning to be signs that philosophical interest is on the upswing again. For instance, Rubio argues that deep learning architectures are compatible with a version of computational functionalism.[89] If this conclusion is correct, then it would be reassuring. Also of interest is a recent paper by Buckner. In this paper Buckner argues that a class of systems called ‘Deep Convolutional Neural Networks’ (DCNNs) seem to perform surprisingly well in various domains. He also suggests that DCNNs can help address longstanding problems in empiricist philosophy of mind, by demonstrating ‘transformational abstraction’. In particular, Buckner is able to propose a model based account of the philosophically puzzling conception of abstraction.[90] If this is a trend which continues, then it seems that connectionism still has contributions to make to philosophy. It is also the case that novel ethical conundrums can arise through the use of these technologies.[91]

Deep learning is certainly a useful technology. For instance, it can be used to do things like classify traffic signs.[92] This is a handy ability, for instance, if one wishes to build self-driving cars. It also appears to be a technology that should continue to be of interest to philosophers, at least based upon the early signs. So, a conclusion of ‘watch this space’, seems to be in order.

# 4 Conclusion

Connectionism can be broadly analysed as having three phases, beginning in the 1940s. The first phase was technologically interesting and generated a certain amount of interest in the media, but had a minimal impact on philosophy. It came to a fairly abrupt end, at least according standard narratives, due to the activities of Minsky and Papert. The second phase of connectionism, due to a combination of factors, gave rise to a great deal of philosophical activity and excitement. It seemed to offer a means of challenging many of the received views concerning nativism, atomism, rules, and representations. Unfortunately, due to the black box problem and the intervention of Fodor, Pylysyn and McLaughlin, quite what the connectionist alternative to these views should be was both controversial and unclear. Eventually, philosophical interest began to wane, but still remains present. More modern developments in connectionist research in the third phase of connectionism is beginning to yield further philosophical insights.

**87** See for instance, Mole “Dead Reckoning in the Desert Ant: A Defense of Connectionist Models”, Arnold, Suzuki, and Arita, “Selection for Representation in Higher-Order Adaption”, and Shea, “Representational Development Need Not Be Explicable-By-Content”.
**88** See Charniak, *Introduction to Deep Learning* and Sejnowski, *The Deep Learning Revolution*.
**89** Rubio, “Computational Functionalism for the Deep learning Era”.
**90** Buckner, “Empiricism Without Magic: Transformational Abstraction in Deep Convulutional Neural Networks”.
**91** See for example, Taddeo and Floridi, “The Debate on the Moral Responsibilities of Online Service Providers”.
**92** See Cireşan, Meier, Masci, Schmidhuber, “Multi-Column deep neural network for traffic sign classification”.

It is certainly the case that the increase in computational power and the increasingly sophisticated architectures that became widely available between the 1940s and the present day has played a role in the rise of connectionism. We will just have to wait and see what interesting insights it will yet yield

## Refrences


Abraham, Tara. *Rebel Genius: Warren S. McCulloch's Transdisciplinary Life in Science*, Cambridge, Mass.: M.I.T. Press, 2016.

Aizawa, Ken. "Connectionism and Artificial Intelligence: History and Philosophical Interpretations" *Journal of Experimental and Theoretical Artificial Intelligence*, 4 (1992), 295–313.

Aizawa, Ken. "Representations without Rules, Connectionism and the Syntactic Argument". *Synthese*, 101:3 (1994), 465–492.

Aizawa, Ken. *The Systematicity Arguments*, Boston, Mass.: Kluwer Academic, 2003.

Aizawa, Ken. "Warren McCulloch's Turn to Cybernetics: What Walter Pitts Contributed". *Interdisciplinary Science Review,* 37:3 (2012), 206–217.

Anderson, James A. and Rosenfield, Edward (Eds.) (1998), *Talking Nets: An Oral History of Neural Networks*, Cambridge, Mass.: M.I.T. Press, 1998.

Arnold, Solvi, Suzuki, Reiji, and Arita, Takaya. "Selection for Representation in Higher-Order Adaption", *Minds and Machines*, 25:1 (2015), 73–95.

Bechtel, William. "Contemporary Connectionism: Are the New Parallel Distributed Processing Models Cognitive or Associationist?" *Behaviorism*, 13 (1985), 53–61.

Bechtel, William and Abrahamsen, Adelle. *Connectionism and the Mind: An Introduction to Parallel Processing in Networks*, Cambridge, Mass.: Basil Blackwell, 1991.

Berkeley, Istvan, Dawson, Michael, Medler, David, Schopflocher, Donald and Hornsby, Loraine. „Density Plots of Hidden Unit Activations Reveal Interpretable Bands", *Connection Science*, 7:2 (1995) 167–186.

Berkeley, Istvan. "What the #$*%! Is a Subsymbol?" *Minds and Machines*, 10:1 (2000), 1–13.

Berkeley, Istvan, and Gunay, Cengiz. "Conducting Banding Analysis with Trained Networks of Sigmoid Units", *Connection Science*, 16:2 (2004), 119–128.

Berkeley, Istvan. "What the <0.70, 1.17, 0.99, 1.07> is a Symbol?" *Minds and Machines*, 18 (2008), 93–105.

Berkeley, Istvan, and Raine, Roxanne. (2011), "An Old Fashioned Connectionist Approach to a Cajun Chord Change Problem", *Connection Science*, 23:3 (2011), 209–218,

Boden, Margaret. *Mind as Machine: A History of Cognitive Science*, Oxford: Oxford U.P , 2006.

Browne, Antony. (Ed.) *Neural Network Analysis, Architectures and Applications*, Philadelphia, PA: Institute of Physics, 1997.

Buckner, Cameron. "Empiricism Without Magic: Tranformational Abstraction in Deep Convolutional Neural Networks", *Synthese*, 195:12 (2018), 5339–5372. https://doi.org/10.1007/s11229-018-01949-1.

Bullinaria, John. "Analyzing the Internal Representations of Trained Neural Networks". In *Neural Network Analysis, Architectures and Applications*, edited by Antony Browne, (1997), 3–26.

Chalmers, David. "Syntactic Transformations on Distributed Representations" in *Connection Science*, 2:1–2 (1990), 53–62.

Chalmers, David. "Connectionism and Compositionality: Why Fodor and Pylyshyn were Wrong", in *Philosophical Psychology*, 6:3 (1993), 305–320.

Charniak, Eugene. *Introduction to Deep Learning*, Cambridge Mass.: M.I.T. Press, 2019.

Cireşan Dan, Meier Ueli, Masci, Jonathan, Schmidhuber Jurgen. "Multi-column deep neural network for traffic sign classification". *Neural Networks* 32 (2012), 333–338.

Chomsky, Noam. *Aspects of a Theory of Syntax*, Cambridge, Mass.: M.I.T. Press, 1965.

Churchland, Paul. M. "On the Nature of Theories: A Neurocomputational Perspective" in *Minnesota Studies in the Philosophy of Science* 14 (1989), 59–101.

Churchland, Patricia S. and Sejnowski, Terrence. "Neural Representation and Neural Computation", *Philosophical Perspectives*, 4 (1990), 343–382.

Clark, Andy. *Microcognition: Philosophy, Cognitive Science and Parallel Distributed Processing*, Cambridge, Mass.: M.I.T. Press, 1989.

Clark, Andy. "A review of "Simple Minds" by D. Lloyd, 1989, M.I.T. Press, London", *Connection Science*, 1:4 (1989), 418–421.

Clark, Andy. "Representation, Development and Situated Connectionism", *Connection Science*, 4:3–4 (1992), 171–174.

Clark, Andy. and Lutz, Rudi. (Eds.) *Connectionism in Context (Human Centered Systems)*, New York, NY: Springer-Verlag, 1992.

Davis, Martin. "Two Notions of Implicit Rules", in *Philosophical Perspectives*, 9 (1995), 153–183.

Davis, Stephen. (Ed.) *Connectionism: Theory and Practice*, New York, NY: Oxford U.P. 1992.

Dawson, Michael and Schopflocher, Don. "Modifying the Generalized Delta Rule to Train Networks of Non-Monotonic Processors for Pattern Classification" *Connection Science*, 4 (1992), 19–31.

Dawson, Michael. Medler, David. and Berkeley, Istvan. "PDP Networks Can Provide Models That Are Not Mere Implementations of Classical Theories" *Philosophical Psychology*, 10:1 (1997), 25–40.

Dawson, Michael. *Minds and Machines: Connectionism and Psychological Modeling*, Hoboken, NJ: Wiley-Blackwell, 2008.

Dreyfus, Herbert. and Dreyfus, Stuart. “Making a Mind verses Modeling a Brain: Artificial Intelligence Back at a Branchpoint” in S. Graubard (ed.) *The Artificial Intelligence Debate: False Starts, Real Foundations*, edited by Graubard, Stephen, Cambridge, Mass.: M.I.T. Press, 1988.
Elman, Jeff, Bates, Elizabeth, Johnson, Mark, Karmiloff-Smith, Annette, Parisi, Domenico, and Plunkett, Kim. *Rethinking Innateness: A Connectionist Perspective on Development*, Cambridge, Mass.: M.I.T. Press, 1996.
Fodor, Jerry. *The Language of Thought*, Cambridge, Mass.: Harvard U.P., 1975.
Fodor, Jerry. *The Modularity of Mind*, Cambridge, Mass.: M.I.T. Press, 1983.
Fodor, Jerry. “Discussion: Connectionism and the Problem of Systematicity (continued): Why Smolensky’s Solution Still does not Work.” *Cognition,* 62:1 (1997), 109–119.
Fodor, Jerry and Pylyshyn, Zenon. “Connectionism and Cognitive Architecture: A Critical Analysis”, *Cognition*, 28 (1988), 3–71.
Fodor, Jerry and McLaughlin, Brian. “Connectionism and the Problem of Systematicity: Why Smolensky‘s Solution Doesn‘t Work”, *Cognition*, 35 (1990) , 183–204.
Haugeland, John. *Artificial Intelligence: The Very Idea*, Cambridge, Mass.: M.I.T. Press, 1985.
Heicht- Nielson, Robert. *Neurocomputation*, New York, NY: Addison-Wesley Pub. Co., 1990.
Hopfield, John. “Neural Networks and Physical Systems with Emergent Collective Computational Abilities”, *Proceedings of the National Academy of the Sciences, USA,* 79 (1982), 2554–2558.
Hopfield, John. “Neurons with Graded Response have Collective Computational Properties Like Those of Two-State Neurons”, *Proceedings of the National Academy of Sciences, USA,* 81 (1984), 3088–3092.
Horgan, Terence and Tienson, John. “Spindel Conference 1987: Connectionism and The Philosophy of Mind”*, The Southern Journal of Philosophy*, Vol. XXVI, Supplement, (1987).
Horgan, Terence. and Tienson, John. “Representations Without Rules”, in *Philosophical Topics,* **17 (1989),** 147–174.
Horgan, Terence. and Tienson, John. (Eds.) *Connectionism and The Philosophy of Mind*, New York, NY: Springer, 1991.
Horgan, Terence. and Tienson, John. “Settling into a New Paradigm”, *Connectionism and The Philosophy of Mind*, edited by Horgan, Terence and Tienson, John (1991), 241–260.
Klahr, David, Langley, Pat and Neches, Robert (eds.), *Production System Models of Learning and Development*, Cambridge, Mass.: MIT Press, 1987.
Lloyd, Dan. *Simple Minds*, Cambridge, Mass.: M.I.T. Press, 1989.
Lovecraft, Howard, P. *H. P. Lovecraft, The Fiction, Complete and Unabridged*, New York: Barnes and Noble, 2008.
Marr, David. *Vision: A Computational Approach*, San Francisco, CA: Freeman & Co., 1982.
McClelland, Jay and Rumelhart, David. *Explorations in Parallel Distributed Processing*, Cambridge, Mass.: M.I.T. Press, 1986.
McClelland, Jay, Rumelhart, David and Hinton, Geoffrey. “The Appeal of PDP”. In *Parallel Distributed* Processing edited by Rumelhart, McClelland and the PDP Research Group, (1986), 503–529.
McCloskey, Michael. „Networks and Theories: The Place of Connectionism in Cognitive Science“ in *Psychological Science*, 2:6 (1991), 387–395.
McCulloch, Warren., *Embodiments of Mind*, Cambridge, Mass.: MIT Press, 1965.
McCulloch, Warren. and Pitts, Walter. „A logical calculus of the ideas immanent in nervous activity“. In *Bulletin of Mathematical Biophysics,* 5 (1943), 115–133.
McDonald, Cynthia and McDonald, Graham. (Eds.) *Connectionism: Debates in Psychological Explanation*, (2 vols.), Cambridge, Mass.: Blackwell Publishers, 1995.
Medler, David. “A Brief History of Connectionism” in *Neural Computing Surveys*, 1 (1998), 61–101.
Miller, Alexander, and Wright, Crispin. (Eds.), *Rule-Following and Meaning*, Ithaca, NY: McGill-Queen’s U. P., 2002.
Minsky, Marvin and Papert, Seymour. *Perceptrons: An Introduction to Computational Geometry*, Cambridge, Mass.: M.I.T. Press, 1969.
Mole, Christopher. “Dead Reckoning in the Desert Ant: A Defense of Connectionist Models”, in *Review of Philosophy and Psychology*, 5:2 (2014), 277–290.
Mozer, Mike and Smolensky, Paul. „Using Relevance to Reduce Network Size Automatically“ in *Connection Science*, 1 (1989), 3–16.
Muller, Vincent. (Ed.) *Fundamental Issues in Artificial Intelligence*, New York, NY: Springer International, 2016.
Newell, Allen and Simon, Herbert. “Computer Science as Empirical Inquiry” in *Communications of the ACM*, 19:3 (1976), 113–126.
Newell Allen. “Physical Symbol Systems” *Cognitive Science*, 4 (1980), 135–183.
Pitts, Walter and McCulloch, Warren. “How We Know Universals the Perception of Auditory and Visual Forms”, *Bulletin of Mathematical Biophysics,* 9:3 (1947), 127–147.
Piccinini, Gualtiero. “The First Computational Theory of Mind and Brain: A Close Look at McCulloch and **Pitts**‘s, Logical Calculus of Ideas Immanent in Nervous Activity‘”, *Synthese*, 141:2 (2004), 175–215.
Plunkett, Kim and Elman, Jeffery. *Exercises in Rethinking Innateness: A Handbook for Connectionist Simulations*, Cambridge, Mass.: M.I.T. Press, 1997.
Pollack, Jordan. “Recursive Distributed Representations”, *Artificial Intelligence*, 46 (1990), 77–105.
Pylyshyn, Zenon. *Computation and Cognition: Towards a Foundation of Cognitive Science*, Cambridge, Mass.: MIT Press, 1984.

Ramsey, William, Stitch, Stephen, and Rumelhart, David. (Eds.) *Philosophy and Connectionist Theory*, Hillsdale, NJ: Lawrence Erlbaum Associated, 1991.
Robinson, David. (1992), "Implications of Neural Networks for How We think about Brain Function", in *Behavioural and Brain Sciences*, 15 (1992), 644–655.
Rosenblatt, Frank "The Perceptron: A Probabilistic Model For Information Storage and Organization in the Brain", *Psychological Review*, 65:60 (1958), 386–408.
Rosenblatt, Frank. *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*, Washington, DC: Spartan books, 1962.
Rubio, Ezequiel (in press), "Computational Functionalism for the Deep Learning Era", to appear in *Minds and Machines*.
Rumelhart, David, McClelland, Jay and the PDP Research Group, (1986), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, M.I.T. Press, (Cambridge, MA), (2 vols).
Samet, Jerry, „The Historical Controversies Surrounding Innateness", in *The Stanford Encyclopedia of Philosophy E*dward N. Zalta (Ed.), (2008)
URL = <https://plato.stanford.edu/archives/fall2008/entries/innateness-history/>.
Schlatter, Mark and Aizawa, Ken. "Walter Pitts and „A Logical Calculus"", *Synthese*, 162:2 (2008), 235–250.
Schneider, Walter. "Connectionism: Is it a Paradigm Shift for Psychology?" in *Behavior Research Methods, Instruments & Computers*, 19 (1987), 73–83.
Sejnowski, Terence. *The Deep Learning Revolution*, Cambridge, Mass.: M.I.T. Press, 2018.
Shea, N. "Representational Development Need Not Be Explicable-By-Content", in *Fundamental Issues in Artificial Intelligence*, edited by Vincent Muller, (2016), 221–238.
Smolensky, Paul. "The Constituent Structure of Mental States: A Reply to Fodor and Pylyshyn" in *Southern Journal of Philosophy*, 26 (1987), 137–160.
Smolensky, Paul. "On the Proper Treatment of Connectionism" in *Behavioural and Brain Sciences*, 11 (1988), 1–74.
Smolensky, Paul. "Tensor Product Variable Binding and the Representation of Symbolic Structures in Connectionist Systems" *Artificial Intelligence*, 46 (1990), 159–216.
Taddeo, Mariarosaria. and Floridi, Luciano. "The Debate on the Moral Responsibilities of Online Service Providers" in *Science and Engineering Ethics*, 22:6 (2016), 1575–1603.
Thorndike, Edward. *Fundamentals of Learning*, New York, NY: Teachers College, Columbia University, 1932.
Van Gelder, Tim. "Compositionaility: A Connectionist Variant on a Classical Theme", *Cognitive Science*, 14:3 (1990), 355–384.
Vera, Alonso. and Simon, Herbert. "Reply to Touretzky and Pomerleau: Reconstructing Physical Symbol Systems", *Cognitive Science*, 18 (1994), 355–360.
Walker, Stephen. "A Brief History of Connectionism and Its Psychological Implications", in *Connectionism in Context (Human Centered Systems)*, edited by Clark and Lutz (1992), 123–144.
Waskan, Jonathon. "A Critique of Connectionist Semantics", in *Connection Science*, 13(3) (2001), 277–292.