Connectionism: A Current Approach to Mind in Cognitive Science and Philosophy

Patrick A. O'Connor

Faculty Sponsor: Eric R. Kraemer, Department of Philosophy

ABSTRACT

The rise of cognitive science, encompassing developments in neurobiology, computer science, linguistics, psychology, anthropology, and philosophy of science in the mid-20th century, has dramatically reshaped the way both mind and brain are conceived of and studied in contemporary academic settings. The mind-body dualism characteristic of Descartes and others has lost ground to monistic, materialistic accounts of mind, insofar as "mind" is conceived as either the brain, or subsystems of it, rather than a distinct nonphysical substance. A deeper understanding of neurology has led many philosophers of mind to favor connectionism as an account of what the mind is and how it works. Connectionism is a way of thinking about the mind as sets of nodes and their connections, and mental states are considered activation patterns of these nodes. In the current paper, I explore the details of connectionist models as well as discuss some of its strengths and weaknesses compared to other accounts of mind in cognitive science. Lastly, I take a look at some purported philosophical implications of this view of mind.

INTRODUCTION

Several questions have guided this research. They are: 1) What are these models and how do they work? 2) Are they actually better than other accounts, as so many of their supporters claim? 3) In what ways do they differ from classical serial processing models? And 4) are philosophical claims about the mind-body relation based on connectionist models justified?

The thesis of which this paper is a part contains a historical sketch of connectionism found in thinkers such as Locke, Hume, Mill, and James[1], which will be left out here due to length considerations. The first main section is devoted to discussing what PDP models are and how they work. The question of how they measure up to the classical serial model is considered, as well as some dialogue between the two camps. Continuing the discussion I raise some deeper questions pertaining to the real differences between the models and the relative merits of each.

This groundwork establishes an appropriate space in which to present P. M. Churchland's biologically driven version of connectionism, his views about why it is the best theory so far, and some of its implications for understanding mind as a connectionist network.

CONNECTIONISM

Brief Description of Classical Cognitive Science

This section identifies the historical context leading to and facilitating the development of connectionism, as well as a little background on "classical" cognitive science.[2]

In the mid 20th century cognitive researchers began thinking in computational terms, and the digital computer seemed like the most accurate type of model available. The computer metaphor has been dominant for over twenty-five years because it is useful as a research tool and believed by many to be the best type of model available. Thus, this approach is termed "classical cognitive science."

Digital computers, acting serially, use programs (i.e. rules) to control their processes that act on data (i.e. symbols) according to the program. Cognition for this approach is serial, with memory storage separate from the main processor, and governed by rules of some kind. The view also assumes that cognitive processing is sensitive to the structure of representations (content), and many mental representations have syntactic structure. A final important principle is that changes in human cognitive states are tractably computable.[3] Another name for this view is the *rules and representations* notion of mentality.

The main classical perspectives[4] assume those three basic conditions. Since speculations about mental activity have perennially invoked internal representations of some kind[5], the computer model organized the study of mind and veered it away from behaviorism. Additionally, there are a couple mistaken notions of the classical model. It is not necessarily deterministic (randomization may be used) and the rules may not be represented in or by the cognitive system itself.

By the late 1970's to 1980's, however, limitations were apparent. The metaphor is unsound given the assumptions of the model. Main problems include integrating new information into the system (e.g. "Mary has left the room"), forming beliefs when there is diverse or contradictory evidence, using memory to solve a problem, deciding what to do, deciding what is moral, and in general making plans. These are broadly termed *the frame* problem. The solution, according to Fodor, is framing all the *relevant* information, called a "frame solution." These fail, though, because human systems are open-ended as to what they can represent, and relevance may be achieved between *any two representations*. In short, the enormous flexibility present in human thought is not captured by the classical digital computer model.

There do not seem to be any ways of handling this problem in the traditional framework. It may be possible that no rule system could accommodate the frame problem. Jerry Fodor has commented,

The problem… is to get the structure of the entire belief system to bear on individual occasions of belief fixation. We have, to put it bluntly, no computational formalisms that show us how to do this, and we have no idea how such formalisms might be developed… In this respect, cognitive science hasn't even *started*; we are literally no farther advanced than we were in the darkest days of behaviorism.[6]

In the wake of these difficulties, connectionism, a brain-inspired model, arose again in the 1980's as an alternative, even though similar attempts had been made earlier.[7] In 1981, though, the approach became mainstream with Hinton and Anderson's *Parallel Models of Associative Memory* and McClelland and Rumelhart's word perception model. The latter's 1986 publication, a two-volume collection, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition,* helped to revive the model as a viable research tool.

What is Connectionism?

The term "connectionism" is really an umbrella term for both PDP (parallel distributed processing) systems and neural network simulations. Nonetheless, there are defining features.[8] Any connectionist model utilizes a general scheme involving a large group(s) of basic units (nodes) which form the processor. Each unit receives and transmits input from and to other units, respectively, the input being excitatory (increases activity) or inhibitory (decreases activity). Relatively simple activation functions determine a node's response to a given input.

An important feature of a connectionist system is nonlinear response, which means that it is capable of binary-style behavior found in classical models, but also graded and continuous response (like neurons, which may be more or less active but typically have threshold levels where "firing" occurs, resulting in activity of a greater magnitude).

Furthermore, what the system "knows" or represents is contained within its patterns of connections and influenced by their weights. Weights are quantifiers that alter activation levels. Since each node has at least one variable weight, maximum flexibility is built into the processor.

Finally, the representations being used are not symbols as in the classical model, but rather the system's own activation patterns across different units. Presenting a word would entail presenting a pattern of activation that corresponds to it. Thus, it would seem that in this approach the brain nowhere contains words, sentences, or rules of logic, as the classical approach assumes.

How the weights, and thus the overall activation, are determined is an important question. Once done manually (math equations), now programs exist which utilize learning algorithms so that the network teaches itself. The system learns by gradually adjusting it's weights in order to approximate the input pattern. Many other types of learning are possible as well, such as pattern detection.

An important question now arises – are there advantages to the connectionist approach, such as overcoming the frame problem which hindered the classical model? Is connectionism limited in any major ways, such that the sentence-logic approach or perhaps an altogether new perspective is preferable? If indeed the connectionist simulations model the brain better, then they should be able to perform and learn as humans do, in situations where the classical digital computer model fails. The next section turns to these questions.

Connectionism vs. Serial Processing: Issues of Debate

In the previous two sections, I showed an initial weakness of the classical digital computer view and laid groundwork for the workings of connectionist systems. Some differences between the two models were also noted. This section is devoted to examining the practical efficacy of each approach – how well can they simulate human

abilities we are most interested in, and what insights can we gather from them regarding how human and animal cognition works?

Horgan and Tienson[9] reveal the complexity and diversity of possible connectionist models, which they use to support their contention that yes, connectionism can deal with the relevance issue already discussed.

An important issue is the sense in which connectionist models employ rules. Is there a significant sense in which they do, and if so, are they different than those of the classical model?

Since the units must have *activation functions*, it is clear that connectionist systems do have rules, but they are not representational rules. The models represent strictly in the form of activated *patterns* of units, thus the rule-driven individual units do not represent anything by themselves.

The system is *nondeterministic* at the representational level even though the units that form the representation are determined. In this sense, "deterministic" means that individual units only have one way of activating at a given time, since their own weight and other units determine what they do. Representational states (collective patterns of activated units) are not determined in this sense – what cognitive states they realize depend on how the activation patterns are set up.

Defeasibility is possible. It is better to think of the system as producing tendencies, or cognitive forces, which are highly complex in that there can be many at a time pushing in different actions. This means stronger forces can override weaker ones. When a new piece of knowledge enters into the system, altering the weights, the relevant states automatically obtain and cognitive forces generate to accommodate it. The advantage over classicism is that the new activation pattern (Sally leaves the room) will inevitably interact with not only other active states, but also the learned nonactive patterns (weights). Thus, rules to cover exceptions in advance are unnecessary since the representational states are not themselves pre-determined.

Besides better solving the frame problem, what other advantages might connectionism offer? McClelland, Rumelhart, and Hinton[10] ask the question, why are people so much better than computers at such tasks as perceiving objects in natural scenes, noting their relations, understanding language, using relevant information in new settings, and making plans? While they acknowledge "software", their response is that it is the brain's unique architecture that

> …is more suited to deal with a central aspect of the natural information processing tasks that people are so good at. We will show through examples that these tasks generally require the simultaneous consideration of many pieces of information or constraints.[11]

Some connectionist-related findings challenge traditional classically postulated bottom-up processing, especially with word recognition. Context is known to play a role in identification and recognition of letters and barely audible words. This suggests that processing is not strictly a serial bottom-to-top affair going from perceptual, to representational, then to interpretive, but rather many different aspects of the stimulus are processed in parallel. In other examples like these, "higher-level" processing, such as knowledge of a context, affected "lower-level" basic processing.

However, in other areas, such as recursion, the model failed to succeed, and Fodor and Pylyshyn have focused on this weakness, especially in "Connectionism and Cognitive Architecture: A Critical Analysis."[12] New types of models were constructed to meet these challenges, including recurrent networks, which have "hidden" units that feed back on themselves, providing additional information for the system to use in dealing with semantic and syntactic learning tasks.

Potential and actual applications of connectionist models are far-reaching, including brain damage, reading, development, pattern completion, and philosophy. PDP systems have many advantages, most notably their added flexibility in dealing with patterns, which may underlie more complex types of behaviors. However, there are several serious worries about the classical and connectionist approaches as well. These include the nonrealistic nature of the "disembodied intelligence" found in computers, the fact that the systems are not intrinsically goal driven but rather programmed, and the idea that cognition is at least partly social, something that artificial models lack.

Further concerns regarding connectionism are the relation of its rule-based units to its representational and cognitive levels and its lack of sentention. If it lacks linguistic representation, then how is all of our language possible? Bechtel[13] lists several differences in task-based performance. Classical models excel at reasoning, but connectionist models are superior in pattern recognition, motor control, and relevant-content memory.

This ambivalence leads him to consider a way of reconciling the two positions.[14] He considers two different ways of understanding this integration, both plausible. First is that pattern recognition may be more basic, or start a cognitive chain going, but also may activate rule based, syntactically represented serial operations. For example, a

connectionist activation pattern may cause you to glance at your gas gauge every so often, in a patterned or regular way, but realizing you need more gas may involve a type of internal process such as "if the needle is on E, then I need more gas. I perceive the needle's on E, I need to go get gas."

Another possibility is the view that PDP systems model the microstructure of cognition, whereas classical models reflect the macrostructure. This "dual mechanism" view is a good middle of the road position, since it would account for the fact that both models excel at certain tasks.

One could still view connectionism, however, as being a viable candidate on its own, especially because of its ability to adapt to unpredictable and novel experiences. It is also difficult to imagine a serial model engaging in active sports, like tennis, for example. Those activities are better described by activation patterns which become strongly weighted for those particular movements. It is hard to conceive of complex fluid movements being governed by predetermined logical rules. Since connectionism implies that there aren't rules for things like that, but rather indeterminate tendencies that are defeasible, connectionism deserves further investigation.

ACTIVATION VECTORS AND ELIMINATIVE MATERIALISM

Introduction

Paul Churchland, a proponent of a biologically biased version of connectionism, offers an interesting philosophical position along with it – he proposes that his form of connectionism entitles him to seriously doubt the efficacy of commonsense psychology.

I present this part of the paper because Churchland attempts to do some interesting things with connectionism, presenting a perspective that is deserving of notice because it calls into question widely held beliefs regarding human mentality. In the following sections I examine his neurocomputational theory as well as his arguments for eliminative materialism, followed by some concerns that I feel he needs to clear up, especially regarding the logistics of connectionism in general.

Churchland's Connectionism

In his essay "Folk Psychology,"[15] he describes his version of connectionism in the following terms: the basic cognitive units are not "sentence-like" but are rather neuronal activation vectors. Activation vectors are patterns of excitation levels across large neuron populations.[16] Statistics equations are used to manipulate these values.

He introduces a related concept as well: vector-to-vector transformation. This refers to the way in which the patterns of activity change as they travel in the brain and body. For example, photoreceptors conduct electrical charges through the eyes down the optic nerve, into the brain, and this is transformed and combined along with other activation patterns and eventually translates into motor neuron activity, bringing about movement in interaction with the stimulus[17].

These patterns of activation levels are finely tuned, and make up the stored information. For Churchland, then, learning does not mean constructing sentences, but rather consists of neural networks altering their connection strengths (weights) and activity levels in order to better transform incoming information into higher-level vectors. At these higher (more complex) levels, there are vectors amplifying and interpreting the information, forming the brain's perceptual and explanatory categories.[18]

Churchland emphasizes the non-sentential element of vector code representations and uses the example of a television to convey the concept. He claims that a specific visual representation on the screen, like on the retina, has no logical or linguistic structure, but is rather a specific pattern of activation or brightness levels across a large population of tiny pixels. He claims that all of our senses code information this way.[19]

The vector coding view is primarily informed by developments in neuroscience, in addition to the cognitive models discussed above. Churchland also advances several reasons for saying that his view is better than both classical models as well as other less neurologically-related versions of connectionism.

The main argument at the beginning of *The Engine of Reason, Seat of the Soul* involves four basic "facts", designed to build a substantial case for his view.[20]

Brute force – parallel systems are faster, and can perform millions of individual computations simultaneously.

Fault Tolerance – parts can malfunction, break, or die in a parallel system and the whole is relatively unaffected, unlike a serial processor.

Mimicking human cognition – the simulations can resolve incomplete sensory information, see complex analogies, focus attention, try many "takes" on a situation, and recognize subtle stimuli.

"Historical Parallels" – Despite seeming wrong, theories and interpretations of experience that clash with common sense have been supportable with the proper insights.

"Combinatorially explosive powers of representation" – vector coding allows enormous possibilities in pattern formation, which translates into increased sensitivity to stimuli.

The model can account for explanatory understanding, prediction, and motor control.

The model helps explain why the brain is built the way it is.

It relieves commonsense psychology of the responsibility of having sets of quantified sentences and a deductive functional structure. On this view, folk psychology is a set of tuned vectors.[21]

His arguments cited thus far favoring connectionism are varied, and Churchland also applies this way of understanding the brain to several research domains, including sensory coding (taste, color, etc.), face recognition, stereoscopic vision, recurrent networks, the social world, and brain malfunction and pathology.[22] While these are interesting in and of themselves, I am more interested in looking at what he takes all this to mean philosophically. The next section, then, is devoted to his argument for eliminative materialism, the view of mind he favors given his scientific perspective.

Philosophical Implications

Churchland states that from two premises – that folk psychology is a theory, and a bad one – he is strongly inclined to drop most of what is considered "mental" in favor of a more powerful, far-reaching neuroscientific account.[23]

His main evidence for this seems to come from the moderate success of connectionism in various psychological domains. He asserts, "Recent science already suggests that Jones's linguaformal theory – folk psychology – fails utterly to capture the *basic* kinematics and dynamics of human and animal cognition."[24]

Support for eliminativist conclusions may be found in the vector coding model's lack of sentential representation and absence of logical rules. However, these conditions are not set in stone. Representation in connectionist models involves some interesting puzzles, such as how it can be determined and rule-governed sub-representationally while at the same time non-rule governed and possibly indeterminate at the representational level.

In the background, however, is a critique of "theory" in general, or more specifically, the logical empiricist view of theory. He challenges their assumptions that theories come in linguistic form, consisting of logical sentences, and that a theory "explains" by logically deducing the explanandum from "rules".[25]

He objects by appealing to theory that proposes that brains most likely do not operate this way. Therefore, he finds it wise to dismiss this classical idea of what theoretical knowledge is and its relation to psychology. Finally, he says that folk psychology is indeed a theory, however, but of a nature which is yet to be defined by other accounts.[26]

These claims taken together become somewhat confusing, but the view is more clearly stated in *Matter and Consciousness*. The basic assumptions are that all perception is 1)theory laden, 2)influenced by the assumptions existing in its background, and 3) always "speculative interpretation. This, at least, is the claim of more recently developed versions of empiricism."[27] He is referring primarily to Quine, who upholds, along with Duhem, that

> …it is the system as a whole that is keyed to experience. It is taught by exploitation of its heterogeneous and sporadic links with experience, and it stands or falls, is retained or modified, according as it continues to serve us well or ill in the face of continuing experience.[28]

If things like truth and reference must be sacrificed for the sake of scientific parsimony so be it.

Churchland does not discuss the issues involved with connectionism, and seems to be overly exaggerating its efficacy. No one has yet proved that it can handle all aspects of human performance on its own. Other connectionist theorists, such as Horgan and Tienson[29], who keep beliefs and desires in their theory, must also be considered. And still other connectionists uphold a view that their model is a microstructure and that classical models serve as a macrostructure.

There is a sense in which Churchland is appealing to a variation of behaviorism. In other words, everything significant about a person would be describable in terms of observable objective entities (at least in principal – we would need ultra-advanced fMRI scans or some other imaging tool).

It becomes difficult to claim a decisive objection if we grant him the two replies to a couple common arguments against eliminativism he presents in *Matter and Consciousness*[30]. The first argument is that introspection proves the existence of mental states. Here he will respond by pointing out "all observation occurs within some system of concepts, and our observation judgments are only as good as the conceptual framework in which they are expressed."[31] The second argument attempts to show that eliminative materialism's being true would mean it was false because to be true it would have to be meaningful, and to be meaningful it would have to have mental components. This argument, however, is question begging, since meaning is precisely what is at issue.

At this point in the discussion, you might want to consider all the advantages that go along with eliminativism. McCauley lists several. [32] There are no worries about how mind-brain identities would work out since there is nothing left to identify with the brain. No category mistakes persist either, since troublesome categories leave with their theories. Science would be liberated from common sense, which has been troublesome for theoretic advancement in the past. Finally, followers of eliminative materialism who are scientific realists can safely depart from the antirealist views developed by Rorty and Feyerabend.

Ultimately, however, Churchland has to show how a theory like activation vectors will gain more explanatory power than traditional efforts in order to win his two premises. This seems implausible, however, and he admits this in *The Engine of Reason, Seat of the Soul:*

> But in one fundamental respect you should rest assured. As will be explained in chapter 5, your physical brain is far too complex and mercurial for its behavior to be predicted in any but the broadest outlines or for any but the shortest distances into the future. Faced with the extraordinary dynamical features of a functioning brain, no device constructible in this universe could ever predict your behavior, or your thoughts, with anything more than merely statistical success.[33]

ASSESSMENT: BENEFITS AND LIMITATIONS

While research with connectionist models in domains such as psychology, artificial intelligence, neuroscience, and philosophy is currently plentiful and productive, limitations must also be acknowledged. While connectionist systems offer an exciting, fresh, and prima facie more realistic way of theorizing about the mind-brain, conclusions regarding the larger philosophical picture must be constrained to fit the evidence. As this paper clearly shows, connectionism is not wholly understood, at both strict empirical and philosophical levels. What "meaning" and "rules" are for connectionist models is still an open question, which further research must explore. Throughout this research I have become very well acquainted with a standard argument in the eliminativist and reductionist literature, namely that all the hard questions will eventually be solved through a successful neuroscience, which will show exactly how the mind relates to the brain. This position seems dogmatic, however. Clearly connectionist models have had success in various domains, but they also have a fair share of philosophical problems. Thinking of mind from a strictly neuroscientific perspective at present does not yield the same kinds of results as the mental talk we are accustomed to when issues such as human behavior are considered. This is not to rule out neuroscience, however, but rather to insist that philosophical conclusions based on cognitive and neuroscientific models must be made with caution at best.

ACKNOWLEDGMENTS

REFERENCE NOTES

[1] See chapters 35 through 39 in *The Philosophy of Mind: Classical Problems/Contemporary Issues*, edited by Brian Beakley & Peter Ludlow (Cambridge, MA: MIT Press, 1994).

[2] Unless otherwise noted, this discussion is drawn from Jeffrey L. Elman, "Connectionism, Artificial Life, and Dynamical Systems," in *A Companion to Cognitive Science, eds*. Bechtel & Graham (Oxford, UK: Blackwell Publishers, 1999) 488-504; and Terence Horgan & John Tienson, "Rules" in the same publication, 660-670.

[3] That an output is computable from an input in a finite number of elementary operations (mathematical computability) *and* in a physical device.

[4] Newell and Simon's physical symbol system hypothesis, Lindsay and Norman's human information processing, and Chomsky's generative linguistics.

[5] Robert S. Stufflebeam, "Representation and computation," in *A Companion to Coginitive Science,* 636.

[6] Jerry A. Fodor, *The Modularity of Mind*: *An Essay on Faculty Psychology*, (Cambridge, MA: MIT Press, 1983), 128-129.

[7] Hebb, Rosenblatt, & Selfridge in the late 1940's and 50's; Cowan, a mathematician in the 1960's; computer science, such as Kohonen in the 1970's, to name a few.

[8] Unless otherwise noted, the sources of the information in this section are given in note 2.

[9] See note 2.

[10] James L. McClelland, David E. Rumelhart, and Geoffrey E. Hinton, "The Appeal of Parallel Distributed Processing", in *The Philosophy of Mind: Classical Problems/Contemporary Issues*, edited by Brian Beakley and Peter Ludlow (Cambridge, MA: MIT Press, 1994)*, 269-288.*

[11] Ibid.,, 269.

[12] Jerry A. Fodor & Zenon W. Pylyshyn, "Connectionism and Cognitive Architecture: A Critical Analysis," in *The Philosophy of Mind: Classical Problems/Contemporary Issues*, edited by Brian Beakley & Peter Ludlow (Cambridge, MA: MIT Press, 1994), 289-324.

[13] William Bechtel, "Connectionism and the Philosophy of Mind: An Overview", in *Mind and Cognition*, ed. William G. Lycan, (MA: Blackwell, 1995) 252-273.

[14] Ibid., 262-264.

[15] Paul M. Churchland, "Folk Psychology," in *On the Contrary*, authors/eds. Paul M. Churchland & Patricia S. Churchland (Cambridge, MA: The MIT Press, 1998), 3-16.

[16] Ibid., 10.

[17] Ibid., 13-14.

[18] Ibid., 15.

[19] Ibid, 13.

[20] Paul M. Churchland, *The Engine of Reason, the Seat of the Soul (Cambridge, MA: MIT Press, 1999),* 11 18.

[21] Arguments 5-8 are found in Churchland, *"Folk Psychology",* 15.

[22] Churchland, *Engine of Reason*.

[23] This claim appears in all of P. M. Churchland's publications cited thus far.

[24] Churchland, "Folk Psychology", 10.

[25] Ibid., 12.

[26] Ibid., 12.

[27] Paul M. Churchland, *Matter and Consciousness* (Cambridge, MA: MIT Press, 1996), 79.

[28] W.V. Quine, "On Mental Entities", from *Proceedings* of the American Academy of Arts and Sciences, LXXX (Boston, 1951).

[29] Horgan and Tienson, "Rules", 668-669.

[30] Churchland, *Matter and Consciousness*, 43-50.

[31] Churchland, *Matter and Consciousness,* 47.

[32] Robert N. McCauley, "Intertheoretic Relations" in *Folk Psychology and the philosophy of Mind*, eds. Scott M. Christensen & Dale R. Turner, (Hillsdale, NJ: Lawrence Erlbaum Associates, 1993), 65-68.

[33] Churchland, *Engine*, 3.