Res Cogitans

Volume 1 | Issue 1 Article 14

7-26-2010

Two Sorts of Connectionist Models: A Critical Analysis

Jack Coughlin
University of Washington

Res Cogitans (2010) 1:113-118 2155-4838 | commons.pacificu.edu/rescogitans

Two Sorts of Connectionist Models: A Critical Analysis

Jack Coughlin
University of Washington

Published online: 26 July 2010

The two main paradigms in cognitive science are computationalism and connectionism. The former takes as its starting point the analogy of the mind as a computer, and so attempts to describe cognition in terms of discrete symbol manipulation, in the manner of a digital computer. Connectionism is a more "bottom-up" approach to cognitive science, assuming the neurological structure of the brain to be fundamentally important. It is the task of this paper to demonstrate that connectionist theories of cognitive science face a dilemma: either they are mere descriptions of the physical implementation of a computationalist system, or they fail to predict and account for an apparently crucial and fundamental feature of human cognition, the systematicity of thought.

First, it is necessary to outline the respects in which connectionism and computationalism differ in more detail. Computationalism holds that cognition occurs at the level of symbols, that it is syntax that does the "heavy lifting", so to speak, when it comes to cognition. Connectionism is the opposite in this respect: paradigmatic research programs in connectionism have typically demonstrated how cognition could occur *without* the use of symbols or rules, instead utilizing base-level connections and node activations to describe thought processes.[1] Connectionist models are parallels to neurological models, so the nodes in such a model can be thought of as idealizations of neurons.

Connectionism is, at first glance, a major shift away from the traditional Turing Machine approach to cognitive science. In doing without symbols, it appears to be genuinely different from computationalism. But thought is quite clearly a semantic process—there is meaning in our thoughts, and any plausible cognitive account must explain the means by which meaning is represented, either syntactically or physically. The relevant difference, if any, between our two theories will come at the one point where their accounts must intersect—representation. Connectionism, then, must demonstrate that its account of representation is significantly different from that of computationalism for it to be a viable upper-level account of cognition.

If we find that connectionism's account of representation is simply a mirror of a computationalist framework, it will turn out that connectionism is not a distinct theory from computationalism, but is simply an elegant solution to the problem of physical implementation. That is, a truly distinct connectionist model must be one in which the computationalist's symbols have no place. For connectionism to say merely that a model *can* do without symbols is not enough – computationalism does not seek to reify symbols, to show that they "exist" in the same sense as a physical substrate. Indeed, computationalist accounts typically ignore the way that symbols are instantiated, although they must be. This fact does not invalidate the computationalist project. Just as chemistry is not invalidated by the ability of physics to (in theory) explain chemical processes, computationalism is not invalidated by the ability of connectionism to explain in excruciating detail the processes of our brains. Whether such an account would truly be an explanation is open to question. What we seek to show in this paper is that connectionism is not an alternative to the computationalist account of cognition, just as physics is not an alternative to chemistry's account of chemical processes.

There are two ways in which representation could occur in a complex system. They are as follows:

1. Full atomicity of thought processes: cognition and thoughts like "Sam has a full head of hair" are fully undecomposable.

2. Thoughts of the above sort can be decomposed into their component concepts: Sam, to have, hair, and so on.

These two mutually exclusive accounts of representation will be referred to as atomicity and compositionality of thought, respectively. Because compositionality of thought is simply the negation of the atomicity hypothesis, these are the only two options to consider. Both end up failing to distinguish connectionism from computationalism in a substantive way.

The first, atomicity, falls to an objection first laid out by Fodor and Pylyshyn[2], called systematicity. Briefly, systematicity is the following property of thought: one's ability to form new thoughts is intricately related to one's ability to form other thoughts. The ability to form a set of thoughts P, Q, R… is sufficient for the ability to form similar thoughts P`, Q`, R`… where P`… are composed of the same *atomic concepts* as P, Q, R… As an example, Fodor and Pylyshyn consider a (highly simplified) thought of the sort "P&Q". Systematicity says that if we can form this thought, we can form similar thoughts like "P", "Q", and (maybe) "&".

This thesis about the nature of thought is highly plausible. For it is inconceivable that one be able to think "The dog ate the homework" and not be able to form other thoughts about the dog, or the homework. Fodor and Pylyshyn point to the apparent impossibility of a natural cognitive system *not* satisfying systematicity, as evidenced by the strong systematicity of language (which is empirically demonstrable).

Atomicity of representation is incompatible with this systematicity condition, however. Under atomicity, "The dog ate the homework" is a monolithic entity, which cannot be decomposed into its component parts. Compositionality of thought is a necessary condition for this systematicity, because without compositionality, the similar thoughts guaranteed by systematicity *might not even exist*. While an atomic mind might very well be able to think about the dog, this is by no means guaranteed, because the ability to form one atomic thought is not sufficient for the ability to form others.

An argument against an atomic, connectionist account of thought might therefore proceed as follows:

(1) Naturally occurring cognition is systematic, as evidenced by the systematicity of language, and the strong intuitive plausibility of the systematicity hypothesis.

(2) Any account of cognition must take into account how cognition actually happens in nature, and so, by (1), must satisfy systematicity.

(3) A necessary condition for the systematicity of thought is compositionality of representation, so by (2) any account of cognition must take representation to be compositional.

(4) Atomicity of thought is incompatible with compositionality, so by (3) any plausible account of cognition cannot be atomic.

There are two potential objections to this argument. The first is with (1)—the systematicity of language is a strong reason to think that thought, too, is systematic, but this does not constitute an airtight argument. The systematicity of thought is going to be an empirical question. However, we *can* say that the systematicity hypothesis seems far more plausible than the alternative, simply because it is difficult to imagine a consistent psychology that is not systematic.

A second possible objection to the argument takes issue with (3). This possibility is hinted at by Fodor and Pylyshyn in their original paper outlining the systematicity objection to connectionism. The idea is that the connectionist could specify that given any representation of thought a**R**b, the system contains a, **R**, and b as separate representations. That is, the presence of an *atomic* representation of "P&Q" (by virtue of the structure of the system, or just by stipulation) implies the presence of all similar representations needed for systematicity. But atomic representations seem to carry no structural information about their contents, indeed, no information whatsoever that is accessible to the system. So there is no way that systematicity-with-atomicity can be guaranteed structurally.[3]

The connectionist can respond to this by saying that perhaps the pattern of nodal activations that represents the thought a**R**b *does* carry structural information about its "contents" a, **R**, and b. Either these three elements always fire together, so that they are in practice inseparable, or a more cohesive activation pattern for a**R**b only gives "clues" about the information it represents. Both of these are ways that a connectionist model can guarantee that systematicity is satisfied by atomic representation. But in providing such a mechanism, we have sacrificed all semblance of elegance or economy. Such a system would never actually arise.

Compositionality appears to be the only recourse left for the connectionist. Atomicity is invalid because it fails to satisfy systematicity, except in an extremely ad hoc manner, so any account of cognition that does not jettison systematicity outright will have to make do with compositionality. But this does not avail the connectionist anything, because any theory that takes a compositional account of representation will be, fundamentally, computationalist. The reason for this isomorphism is due to the fundamental difference between the two theories: computationalism does the hard work with symbols, while connectionism works at a subsymbolic level.

What would compositionality mean for a connectionist model? It would simply mean that representations can be broken up into identifiable and distinct components at the physical level.[4] Every atomic concept (P, the dog, etc.) is represented by a physical pattern of nodal activations. Compositional representations (full sentences, e.g.) are composed of numerous patterns of activation firing in concert. Each of these representational components ("P", the dog, etc.) will have causal power at the level of implementation, of connections and nodes.[5] So cognition, for the connectionist, is tokening of these "pieces" of representation at the physical level. Already the connectionist model seems to have lost some of its uniqueness – we can point to one pattern of activation and say that the system is thinking about the classic Wilco album *Yankee Hotel Foxtrot*. The idea behind connectionism is, at least partially, that this should not be possible, that inputs go in and outputs come out, and that only the system knows what happens in between. But it is still unclear that what we have is a computationalist system.

For our compositional connectionist system to also be computationalist, symbols representing elements of cognition would have to be able to effect other symbols, also representing elements of cognition. Where do the symbols appear? The answer is that they already have! If we were to draw a (high-level) map of our connectionist network, no one could fault us for simply writing *Yankee Hotel Foxtrot* instead of drawing the whole map of nodes that corresponds to that album, and doing the same for other representations. Such a map would be, for certain purposes, just as useful if not more useful than the more detailed, node-by-node map. The result would be a set of boxed keywords (or symbols), connected by lines of activation, just like a connectionist graph, but at a more abstract level.

Naturally, it is not the case that the symbols can be said to actually have causal power to effect cognition. A physicalist account of cognition will have to admit that causality originates most fundamentally in the physical substrate of cognition. But insofar as we can call reliable correlation causation, it makes sense to speak of symbols "causing" other symbols, in the sense that the one is always and necessarily followed by the other.

This is just what we mean when we speak of a computationalist model. Computationalism seeks to provide a more abstract account of the same phenomena as connectionism. It is to connectionism what chemistry is to physics. If a symbolic account is theoretically useful, then connectionism has not unseated computationalism.

In short, compositionality entails computationalism, because in any system of representation where symbols are free to combine, recombine, and act as part of a wide variety of thoughts, processes at the symbolic level will be just as complex as those at the lower level. There is a symmetry between lower-level physical processes and upper level processes which is only avoided by an atomic account of representation. But since atomicity is incompatible with systematicity, it turns out that a symbolic account of cognition is just as relevant and useful as a neurologically motivated one. Connectionist architecture is merely implementing the computationalist system.

It is seen that connectionism, plausibly construed, can be nothing more than an implementation of computationalist symbol manipulation architecture. Indeed, the only difference between the two theories is that connectionism, at first, seems to be atomic in its account of representation. But this turns out to be unsystematic, and no hope at all for connectionism as a unique and significant contribution to cognitive science.

---

[1] An example is *On Learning the Past Tenses of English Verbs*, by Rumelhart and McClelland. Their system uses no rules in modeling the learning process observed in children for the past tense in English.

[2] Fodor and Pylyshyn, *Connectionism and Cognitive Architecture: A Critical Analysis*

[3] The possibility that the systematicity is *designed* into the system barely merits consideration—obviously once we allow our cognitive theories to include designers, we can make do with any sort of crazy theory.

[4] If we cannot break up representations in this way, we just have atomicity.

[5] This is necessary for systematicity to obtain, that each "piece" of a representation be independent and causally powerful.