Are Connectionist Networks A Proper Response to the Physical Symbol System Hypothesis?

Joshua Herring

Introduction

It is often claimed that connectionist networks provide an alternative model to traditional symbolic approaches of how intelligence might arise. Such claims represent a direct challenge to Newell and Simon's Physical Symbol System Hypothesis - the idea that any intelligent system is *necessarily* a symbol system. The basis of the challenge seems to rest on the distribution of information across the network. Networks lack semantic transparency: their operational parts are not directly associated with the operations of the process or function being represented. Rather, information is said to be *distributed* across the architecture of nodes and connections. This stands in contrast to something like an abacus, which involves the representations of the semantically relevant parts directly in carrying out its operations. However, this explanation of the difference is perhaps not as convincing as it seems at first glance. An argument can be made that connectionist networks do nothing other than instantiate symbolic functions. That is, whether or not the architecture of the network is itself directly symbolic, the function which the network learns to approximate when it "implements cognition" *is*. Further, it's not even clear that there's anything anti-symbolic about the architecture. After all, a network is set up as a series of logical gates like any other. Nodes either receive enough input to fire or they don't - and combinations of such binary circuits make up the whole of the functioning of the system. It seems likely, therefore, that symbolic equivalents of networks can be found *on the level of operation*. It isn't just that networks approximate symbolic functions, in other words, it's that they are, from a certain point of view, hardwired symbolic systems themselves.

However, these objections do not cover the full scope of the debate. Connectionist networks clearly contain at least some subsymbolic levels of representation. The all-or-none output masks these by enforcing firing thresholds, but there are meaningful senses in which the network can be said to capture strength of certainty (in the form of higher or lower levels of activation) about its conclusions across all levels of the process of forming output. More importantly, connectionist networks come with a robust and useful learning algorithm that depends on these subsymbolic levels of representation. Though the final "resting" state of the network may in some sense *be* a symbol system, the level

on which learning takes place is not necessarily.

Fodor and Pylyshyn: the "two-horned" defense

The most common objection to the idea that connectionist networks are an alternative model of cognition is that any network that models cognitive behavior will turn out to be a connectionist implementation of a symbol system. This line has been taken by many, but it is perhaps best articulated (and certainly at its most influential) in [9].

Phillips has characterized this as a "two-horned" approach [11]. People attempting to object that connectionist networks are different in kind from symbol systems will run into the objection that if a connectionist network merely approximates a symbol system then it is failing to capture vital facts about cognition, and people attempting to object that connectionist networks can exhibit the required systematicity encounter the objection that it is then merely implementing a symbol system. Since one of these approaches would seem to be required, it looks as if there's no way to win.

Recent Perspectives

There is something about Fodor and Pylyshyn's approach that seems like cheating. People will say that they have simply stacked the deck in favor of their interpretation: anything which meets the relevant description is a classical symbol system, and by virtue of being so any other model implicated is "merely an implementation" by definition. It seems very much like they have defined as classical "anything which could potentially contribute to our understanding of cognition" and thus cleared the field of competitors. However, this is to attribute too much *prima facie* legitimacy to connectionism as a model of cognition. In fact, Fodor and Pylyshyn are correct that the burden of proof is on the "interloper" to demonstrate that it can capture facts about human cognition that the more established model cannot. So has connectionism in fact demonstrated any such thing?

In fact, it probably has. It is pointed out in [13] (among many others), for example, that connectionist networks are better suited to handling classification and approximation problems. In general, there are a number of results from Psychology that come naturally to connectionist networks. Probably more importantly, it is not clear how symbol systems can handle learning (especially evolutionary learning) without a great deal more innate specification than probably actually exists in the human brain. There are, in short, tasks associated with cognition for which symbol systems fall prey to the same kind of objection that Fodor and Pyyshyn give against connectionsists attempting to approximate systematicity though highly sophisticated assocationism: that attempting to approximate things like gradient descent and exemplar-model classification tasks in a purely symbolic system would require an unmanageable explosion in

the number of required rules and innate specifications. Interestingly, these are precisely the tasks which do not require systematicity in the sense of [9].

Not surprisingly, most recent approaches to this problem do not commit themselves absolutely to either the classicist or the connectionist approach. Rather, there is a move to talk about integrated models. [1] is one such attempt, offering a concrete specification for a method of translating between logic program models and connectionist models. That is, it defines an operator which, when embedded in connectionist networks, allows for a straightforward implementation of logic programs in this medium. While intended primarily as a model for practical application and implemenation of research, it has cognitive implications in terms of explaining how a single architecture can exhibit the appropriate behaviors at the appropriate times (or with regard to the appropriate problems).

References

[1] S. Bader, P. Hitzler, and A. Witzel. Integrating first-order logic programs and connectionist systems — a constructive approach, 2005.

[2] Sebastian Bader and Pascal Hitzler. Dimensions of neural-symbolic integration - a structured survey.

[3] Peter beim Graben. Incompatible implementations of physical symbol systems. *Mind and Matter*, 2(2):29–51, 2004.

[4] D. J. Chalmers. Why Fodor and Pylyshyn were wrong: The simplest refutation. In *Proceedings of The Twelfth Annual Conference of the Cognitive Science Society, Cambridge, MA, July 1990*, pages 340–347, Hillsdale, NJ, 1990. Lawrence Erlbaum Associates.

[5] Morten H. Christiansen and Nick Chater. Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, 23(2):157–205, 1999.

[6] Morten Hyllekvist Christiansen. Infinite languages, finite minds - connectionism, learning and linguistic structure.

[7] Mark Derthick and David C. Plaut. Is distributed connectionism compatible with the physical symbol system hypothesis? In *Proceedings of the 1986 Cognitive Science Conference*. Lawrence Erlbaum Associates, 1986.

[8] Jerry Fodor and Brian McLaughlin. Connectionism and the problem of systematicity: Why smolensky's solution doesn't work. *Cognition*, 35:183–204, 1990.

[9] Jerry A. Fodor and Zenon W. Pylyshyn. Connectionism and cognitive architecture: a critical analysis. In S. Pinker and J. Mehler, editors, *Connections and Symbols*. MIT Press, Cambridge, Mass., 1988.

[10] P. Hitzler, S. en, H. Anthony, and K. Seda. Logic programs and connectionist networks, 2004.

[11] S. Phillips. Connectionism and the problem of systematicity, 1995.

[12] Paul Smolensky. Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence*, 46:159–216, 1990.

[13] David L. Waltz. Connectionist models: Not just a notational variant, not a panacea. In *Proceedings of the Third Workshop on Theoretical Issues in Natural Language Processing (TINLAP-3)*, pages 58–64. Association for Computational Linguistics, 1987.