

Function, Sufficiently Constrained, Implies Form

Commentary on Green on Connectionist-Explanation

(French, R. M. & Cleeremans, A. (1998) Function, sufficiently constrained, implies form: Commentary on Green on connectionist-explanation. *Psychology* 9(21) psyc.98.9.21.connectionist-explanation.18.french)

Robert M. French (rfrench@ulg.ac.be)
Psychology Department (B33)
University of Liege,
4000 Liege, Belgium

Axel Cleeremans (axcleer@ulb.ac.be)
Séminaire de Recherche en Sciences Cognitives
Université Libre de Bruxelles
1050 Brussels, Belgium

Abstract

Christopher Green's (1998) article is an attack on most current connectionist models of cognition. Our commentary will suggest that there is an essential component missing in his discussion of modeling — namely, the idea that the appropriate level of the model needs to be specified. We will further suggest that the precise form (size, topology, learning rules, etc.) of connectionist networks will fall out as ever more detailed constraints are placed on their function.

1. Every model should be accompanied by a clear indication of the level for which the model is appropriate. This is a necessary condition to making sense of a model of anything. If this is not done -- and, unfortunately, it rarely is -- every model becomes either false or meaningless at some level of consideration. For example, a symbolic, grammar-based model of language processing could undoubtedly be used to do a good job of translating, for example, sewing machine-assembly manuals. Thus, for the high-level task of translating short, declarative sentences that employ a very restricted vocabulary, the proposed model is indistinguishable from other models that might use lower-level structures or more complex mechanisms to achieve the same end. It is only once we begin to probe this level in a Turing-test like manner (Cleeremans & French, 1996; French, 1990, 1995), that differences in the two models will become apparent, thus revealing something about the mechanisms underlying the behavior.
2. The reason that this approach is necessary is that at some level all models of any phenomenon will fail - since the model and the phenomenon being modeled are not literally one and the same thing. While this certainly should not automatically disqualify them from being good models, it does mean, however, that modelers must specify the level at which their models are appropriate. For example, the Newtonian and the Einsteinian models of physics are indistinguishable for the speeds at which most objects generally encountered on earth move. However, when these models are "probed" at speeds approaching that of light, the Newtonian model falls short.
3. In a somewhat analogous manner, we can compare symbolic and connectionist models of cognition. As long as we restrict cognitive behavior to certain high-level phenomena encountered, say, in chess-playing, route-finding, sentence parsing, grammatical transformations, strategy-planning, etc., a symbolic account is, like Newtonian physics, a perfectly adequate (and appropriate) model. However, as soon as we begin "probing" the symbolic account by considering how well it handles generalization, how well it deals with partial information, how fault tolerant it is, etc., insufficiencies in this account come to light. Connectionist modelers believe, by and large,

that the number and magnitude of these insufficiencies justify adopting a subsymbolic level of modeling. But is the new (connectionist) model that does what the old (symbolic) model does, in addition to being able to handle tasks the old model could not, a better model? Not necessarily, Green would say. It depends on the degree to which we can understand what the model is actually doing.

4. This leads us to one of Green's (1998) key points. Consider the problem of designing a system that will be able to balance a free-standing pole on a mobile cart by moving the cart in an appropriate direction each time the pole begins to fall over. One system, a traditional symbolic system, consists of rules based on equations that will specify exactly how the cart must move so as to maintain the pole upright. Studying this system ("opening the box," so to speak) really does allow us to know something about underlying pole-balancing mechanisms. Furthermore, we can make predictions about what would happen if, say, we used a thicker pole, a heavier pole, a longer pole, etc. Compare this to a second system - connectionist this time - that "merely" learns how to move the cart so as to balance the pole. In the latter case, Green would claim we have learned nothing because there is no explicit semantic content to the nodes and connections in the connectionist network. In other words, looking inside the box tells us nothing that we didn't know already.

5. But this is false. Looking inside the connectionist box does indeed tell us a LOT and if we look carefully, we will discover a lot more. To begin with, we learn a highly non-obvious fact: That a system with a particular topology, particular rules of activation passing, certain learning algorithms, etc., can, in fact, produce pole-balancing behavior. That, in itself, is extremely surprising and can only be achieved by a vanishingly small subset of all possible architectures. Furthermore - and Green does not seem to be aware of the research in this area - there are a wide range of techniques to extract high-level rules from neural networks (Towell & Shavlik, 1993). In fact, in cases in which this can be done, one can argue without too much difficulty that the system is following a rule, even if that rule is not implemented as it would have been in a symbolic system. When this is done, groups of units acquire the semantics that boxes and labels have in symbolic models. What makes this particularly exciting in the case of connectionist networks is precisely the fact that the semantics emerge out of the interplay of processing principles specified at a sub-symbolic level. Thus, in contrast to more classical approaches, connectionist theorists believe they can learn something by evolving their networks rather than by fully specifying each of their components (see Content & Frauenfelder, 1996). While this strategy can indeed produce models that merely fit the data, we believe it is just as clear to most connectionists as it is to Green that merely fitting the data does not amount to theory-building. That is precisely why there are so many instances of connectionist research in which the emphasis is put on detailed analysis of internal representations, processing characteristics, and the like. One needs to look no farther than Rosenberg's (1987) analysis of the "structure of NETtalk's internal representations" to realize that ever since the early days of connectionist modeling, researchers were developing ways to peer inside the box and understand what was going on. What is gained in this process of dual exploration (of the modeling and modeled spaces) -- namely, an understanding of function that is rooted in the dynamics of evolution, development and learning -- is unlike anything that can be achieved by specifying the semantics of each component of the model, as in more traditional models.

6. So, what do the nodes of a neural network correspond to? The answer to this question seems to be of overwhelming importance to Green. He believes that they must be made to correspond as closely as possible to real neurons. But why stop at neurons? In fact, this would seem to be a rather arbitrary choice. Why not synapses? Why not vesicles on the membranes of synapses? Why not neurotransmitters? Why not the molecules making up neurotransmitters in the vesicles in the synapses of the neurons? The point is that even if the nodes of connectionist models were designed to correspond as closely as possible to real neurons, they would not be real neurons and,

consequently, the model would be false at some level. Just ask people who do actually model the collective behavior of real neurons. Even though the neurons they use are incomparably more sophisticated than the modified McCulloch-Pitts types used in most connectionist models, these researchers nonetheless come under fire from neurobiologists who continually complain about the egregious oversimplifications of these models.

7. The point is that most researchers in cognitive modeling believe that we will not need to go below a certain level of modeling precision (groups of neurons?) in order to model what is normally thought of as the full range of human cognition. But for the moment it is premature to insist on physical correspondences with real brains. Connectionists are still grappling with issues of function: what types of organizational architectures are able to give rise to phenomena such as priming, implicit learning, incubation, tip-of-the-tongue, gradual forgetting, etc. As our ability to simulate human cognitive function becomes more and more refined, form will invariably follow. Function, sufficiently constrained, implies form.

8. Consider an example to illustrate this crucial principle. Assume the function under considered is the highly unconstrained "to fly". With respect only to the ability to fly -- i.e., moving from point A to point B in the air -- hummingbirds and jet airplanes are indistinguishable. But as we "probe" function in an ever more detailed manner, we claim that form will be increasingly determined. For example, we can probe the flying function with questions like "Does it allow landing on a sunflower?" (Yes.) "Does it permit turning 90 degrees in mid-flight?" (Yes.) "Does it include the ability to hover in mid-air?" (Yes.) "Does it produce a high pitched buzzing sound?" (Yes.), etc. The longer and more specific the probing of the function of flying is, the more the form of the flying object in question will ultimately come to resemble a hummingbird.

9. For this reason, connectionists are justified in remaining vague in their claims of physical correspondences between their networks and the human brain. And it is for this reason that when Green insists that "...the success of connectionist models seems to DEPEND upon the fact that any give unit can send excitatory impulses to some units and inhibitory impulses to others. No neuron in the mammalian brain is known to do this...", his claim falls on deaf ears. Not only has it been clearly been demonstrated (see, for example, Shepard, 1988, p. 163) that a single neuron can have more than one type of neurotransmitter and "can mediate opposite synaptic actions to different follower cells or to a single follower cell," but, more importantly, his comment misses the point. Whether nodes correspond to single neurons, groups of synapses, or groups of neurons is, in some sense, irrelevant. Connectionists are trying to establish overall architectures that implement certain specific principles and allow them to simulate various aspects of human cognitive function. The question of form -- in particular, exactly what nodes of connectionist networks correspond to -- will take care of itself as we put more and more constraints on the networks' function.

ACKNOWLEDGMENTS: This work was supported by grants from the Belgian government: FRFC #2.4605.95 F and IUAP #P/4-19. Axel Cleeremans is a Research Associate with the National Fund for Scientific Research (Belgium).

REFERENCES

- Cleeremans, A. & French, R. (1996). From Chicken Squawking to Cognition: Levels of description and the Computational Approach in Psychology. *Psychologica Belgica* 36(1-2), 5-29.
- Content, A. & U.H. Frauenfelder (1996). On the need for computer modeling: The case of language processing. *Psychologica Belgica* 36(1-2), 113-144.
- French, R. (1995). *The Subtlety of Sameness: A theory and computer model of analogy-making*. Cambridge, MA: MIT Press.
- French, R. (1990). Subcognition and the Limits of the Turing Test. *Mind*, 99(393), 53-65.

- Green, C.D. (1998) Are Connectionist Models Theories of Cognition? PSYCOLOQUY 9 (4)
<ftp://ftp.princeton.edu/pub/harnad/Psycoloquy/1998.volume.9/psyc.98.9.04.connectionist-explanation.1.green>
- Rosenberg, C. (1987). Revealing the Structure of NETtalk's internal representations. *Proceedings of the Ninth Annual CogSci Conference*. 537-554
- Shepard, G. (1988) *Neurobiology*. Oxford: Oxford University Press.
- Towell, G. & Shavlik, J. (1993). The extraction of refined rules from knowledge-based neural networks. *Machine Learning*, 13:1, 71-101.