Psycoloquy

Psycoloquy is a **refereed international, interdisciplinary electronic journal** sponsored by the American Psychological Association (APA)
and indexed by APA's **PsycINFO**
and by Institute for Scientific Information.

Psycoloquy publishes target articles and peer commentary in all areas of psychology as well as cognitive science, neuroscience, behavioral biology, artificial intelligence, robotics/vision, linguistics and philosophy.

Logo Design: Guy Hivroni

---

psycoloquy.98.9.47.connectionist-explanation.31.greco Sat Oct 3 1998

ISSN 1055-0143 (23 paragraphs, 3 notes, 22 references, 477 lines)

PSYCOLOQUY is sponsored by the American Psychological Association (APA)

CONNECTIONIST MODELS CAN REVEAL GOOD ANALOGIES

Commentary on Green on Connectionist-Explanation

Alberto Greco

DISA Psychology Laboratory

University of Genoa

Genova, Italy

greco@disa.unige.it

http://www.lettere.unige.it/sif/strutture/1/home/greco/index.htm

ABSTRACT: Green (1998a) argues that distributed connectionist models are not theories of cognition. This is reasonable if it means that the explanatory role of connectionist models is not clear, but Green's analysis seems directed against the wrong target when he applies a realist position to models. His argument confuses models with objects. Models are useful as long as they establish analogies between unknown and known phenomena; but not

all details are important. The real problem may concern the explanatory role of connectionist models (which is what Green also seems concerned about), but then it should be formulated on different grounds. If they are intended as cognitive models (and not as mere AI artifacts), their internal operations should be describable (by analogy) using a cognitive vocabulary. This is often not the case with connectionist models. Are they always useless as cognitive models then? I cannot share Green's conclusion that the only hope for connectionism is to model brain activity. On the contrary, because the most attractive feature of connectionist models is that they can perform cognitive tasks using no symbols, they can be useful tools for studying (by analogy) the origin and grounding of symbols.

1. Green (1998a) in his target article and subsequent replies (Green 1998b, c, d), argues that distributed connectionist models (henceforth, simply "connectionist models") are not theories of cognition. One of the purported reasons is that their explanatory role is not clear, and that "what one can really be said to have learned" about a cognitive phenomenon is difficult to specify (para. 14). One cannot find this an unreasonable concern, having oneself made a similar epistemological analysis some years ago (Greco, 1990a). It is not clear, however, that Green's analysis is conducted on the right

grounds, and there are problems with his his conclusion that the only hope for connectionism, so far, is to model brain activity.

2. In the present commentary, I will try to show the following: (i) Green has overlooked the fundamental difference between theories and models. (ii) A model is not directly concerned with the object of investigation, but with the object of a different discipline, to which it is analogous. (iii) Connectionist models are misused when cognitive [1] analogies are not clearly explicated. (iv) Beyond symbol tokens identifiable by "post-hoc" statistical analyses, there is a cognitive "reality" to which connectionist processing could be analogous: nonsymbolic and presymbolic activities; these capture important aspects of cognition and cannot be modelled otherwise.

3. Green's main argument is that to consider connectionist models as theories of cognition, in a realist sense, it should be possible to see what each "theoretical entity" refers to. He claims that every single unit or connection in a net should be considered a single theoretical entity (para. 10) and asks what they correspond to "ontologically" in cognitive reality.

4. As earlier commentators (Lee et al., 1998; Goldsmith, 1998) have already noted, this reasoning is flawed by a confusion between a model

and a theory. This distinction ought to be made clear in advance to avoid misunderstandings, but Green never explains it. Yet, his whole argument is based on a contrast between models and theories. Such terms are explicitly contrasted where Green (1998a) states initially that "it is not clear... in what sense such models are to be considered THEORIES [original emphasis] of cognition." In subsequent replies, however, Green (1998b; 1998c) avoids discussing this relationship further, merely claiming that it is an old and unresolved question among epistemologists. He suggests that the "semantic approach" to the theory/model relationship is a "morass" that is best avoided because it involves "as many views of the relationship between models and theories expressed... as there are scientists and philosophers of science who have considered the question" (Green, 1998b, para. 1).

5. This may be true; but then if a full epistemological analysis is not convenient in the current discussion the only option left is to adopt the standard practice accepted (often implicitly) in a particular scientific field. As Green fails to explain fully what he means in using these two terms, we are entitled to suppose that his view coincides with the most common one. Green should concede that, in current practice: (i) cognitive scientists use general statements ("theories") about cognitive phenomena; (ii) they often also try to

explain such phenomena by using computational systems ("models") which are effective in performing cognitive tasks.

6. Green's assumption is that every single unit or connection in a net should be considered a single theoretical entity (para. 10): taken this way, clearly no unit or connection is essential; nor does it correspond to any theoretically relevant term. This assumption seems untenable. The main counter-argument is already there in the target article, but rejected: one can say that it is not single units, but "only a network with a certain general sort of architecture and certain sorts of activation and learning rules" that is being tested (para. 11). Green rejects this possibility on the grounds that it would not be "a bold conjecture," and because to move away from the details of a theory would mean "to shield them from the possibility of refutation."

7. Only one thing enables us to describe or explain cognition by using systems (models) from another domain: analogy. This point has been already made in the present discussion, but Green does not seem receptive to it. He quickly dismisses it with a generic statement: "[the] claim that 'the theory specifies the similarity relations that obtain... between the model and the phenomenon to be explained' is a well-worn one, but not one that has historically been able to hold much water" (Green, 1998b, para. 1). I will accordingly try to put it more

explicitly. It is important to realize that a model (e.g., a working network) is a phenomenon in its own right that can be described and explained by a specific theory (in our case, the connectionist theory).
If one is able to see any similarity between two phenomena, one can use the one of the two that is better understood, more manipulable, and simpler, to describe or explain the other. Neural nets are simpler, more manipulable and (perhaps) better understood than human cognitive systems. Not all their details are important, however.
When analogies are made between a known phenomenon (the explanans) and an unknown one (the explanandum), a theory must specify which of the details are important and which are not ("positive" vs. "negative" and "neutral" analogies, as Hesse, 1966, first called them).

8. So why insist on focusing on the details of single units? In connectionist theories, single units or connections are not important, whereas architectures or patterns of activation (or other aspects) might well be. Green asks what units and activation rules refer or correspond to, in reality, but his argument incorrectly demands a "reference" relationship in models, whereas a "similarity" relation should be appropriate. They do indeed refer to "something real" happening in the net, but this can be described and explained only by the connectionist theory. When they are used as models of something else, then analogy is what gives them scientific value, just as in

other kinds of modelling, where mechanical or electrical systems model real physical phenomena. Physics, as a theory, can properly describe and explain these. When such systems are used as cognitive models, they do not "refer" to any cognitive reality at all. The question is whether or not they are analogous to cognitive phenomena, not whether they correspond to them in every detail.

9. Green's ontological question about the "reality" of units, connections, and other connectionist properties, then, involves a confusion between models and the modelled reality. To be bent on seeking an "ontological" correspondence between nodes and "something real" in the cognitive domain is like asking for a correspondence between "real" cognitive variables and ALL variables used in a programming language (e.g. Lisp or Prolog in traditional-style simulations), or even their ad hoc implementations in assembler routines. These are irrelevant details for certain purposes. Green's examples -- genes, atoms, even memory stores -- are not models [3] but can be considered the direct and proper objects of investigation. In contrast, nets are not the reality under investigation, but a model of it; the right question then concerns their "correspondence," by analogy, to reality.

10. Green does raise a genuine problem, however, even if he directs his

analysis against the wrong target. The real problem is that, in using any model (connectionist or not), one expects to learn MORE about a cognitive phenomenon. "To know more" may mean different things, some of which have more scientific value than others: to explain, to predict, or "simply" to acquire new ("heuristic") ideas. I am unsure that all connectionist models offer such benefits, but their shortcomings are not the ones that Green attacks in his target article.

11. Nets belong to a special category of models: simulations. The aim of a simulation is to reproduce a phenomenon -- not exactly, but within the above-mentioned limits of all models: in a simplified way, by identifying crucial analogies. A net is treated as an experimental subject and a simulation is treated as an experiment, with stimuli presented to the net and responses generated by it. Nets are then described using the vocabulary of cognition.

12. The choice of a vocabulary is not irrelevant when one passes from commonsense to science. Disciplines differ from one another because each accepts as relevant only some descriptions of the common (prescientific) reality. Consider a simple phenomenon such as "rising temperature": This can be described from a variety of disciplinary viewpoints -- physical, meteorological, even psychological -- according to which of its aspects are considered relevant or what relations are

established. Cognitive science is concerned with all aspects of knowledge and its accepted vocabulary includes stimuli, learning, problem solving, etc. Nets seem perfect examples of cognitive models, with input unit activations described as stimuli, outputs as responses; the overall operation is described as learning, perception, categorization, etc. What enables us to use such a vocabulary is again analogy. If one described a network operation only with: "activation spreads thus and so among units," it would be useless.

13. Some connectionists claim that this is enough, because if one is able to reproduce a phenomenon, one can also understand it. It is easy to concede that this understanding could consist in arriving at new ideas (heuristic value), because the new ideas will presumably be about cognitive or psychological matters. But we should be more cautious in claiming that this kind of understanding consists in explanation. A simple reproduction is not eo ipso an explanation. To explain means to give reasons, to indicate why things happen. This can be done in various ways (see Nagel, 1961): by considering a phenomenon as a particular case in a class of general facts (hypothetico-deductive explanation), or by showing what function the phenomenon has in a context, or how a previous event generates a subsequent one (genetic explanation). In any case, explanation is impossible if the vocabulary of a different discipline is used: one cannot explain an electrical

phenomenon by using the vocabulary of economy, speaking about the market-value of different wires. What one said might all be true, but it would be irrelevant. One might well be able to use chemical vocabulary, but only if correspondences between electrical and chemical properties are shown.

14. The problem with connectionism is that the analogy should hold not only when describing but also when explaining. I would rephrase Green's call for an "obvious path to follow to get from the 'high level' of behavior and cognition to the 'low level' of units and connections" (Green 1998a, para. 22), asking instead for an analogy-based path. One should be able to say how outputs (interpreted as responses) are produced from inputs (stimuli) by using a corresponding cognitive vocabulary. The analogy should hold for all model parts: inputs, outputs, and internal patterns. Connectionists, however, sometimes shift unexpectedly to a different language -- about activations, cell organization, trajectories, tensor products, etc. -- and all analogies are lost along the way. If connectionist models fail to show analogies in the crucial points that should be explained (viz., how to go from input to output), they are useless as cognitive models (even if they work, in which case they may be good AI artifacts).

15. What conclusion follows from all this? Certainly not Green's conclusion that the only hope for connectionism to date is to model biological neural activity. If the analogy between nodes and neurons (or between activations and neural processes, etc.) were to prove well grounded, that would be a neural simulation, which would presumably use the vocabulary of neuroscience, not that of cognition. It would then be up to neuroscientists whether to accept it as a good simulation (and probably most would not). Is our only conclusion, then, that connectionist models are not suitable as cognitive models? Perhaps we should not put all models in the same basket. The last part of this commentary will be devoted to discussing two important requirements which make the difference: that the cognitive analogies should be shown explicitly and that good analogies do exist. With respect to the second topic, I shall suggest some aspects of cognition which can be captured only by connectionist nets.

16. The most common way to show the cognitive analogies is to analyze a net's internal representations after it has mastered a task. Many connectionists claim that post-hoc statistical analyses can accomplish this (see some of commentaries in this discussion: Medler & Dawson, 1998; Lee et al., 1998); others (Green included) are more critical. Do statistical analyses really help make analogies explicit?

The answer may be affirmative, but a possible confusion should be

clarified here: A simulation can be either constructed or used. If it is constructed, analogies must be clear and explicit from the beginning; if used, analogies are not immediate, but come from the interpretation.

17. Traditional cognitive models are usually "constructed" on the basis of the description of a cognitive process. By contrast, connectionist models seem to be used as models; this means that with them the simulation method is reversed. Normally, the first step in modelling should be a theoretical or empirical analysis of the process under investigation; the natural outcome of this analysis is a representation of relevant variables, which may be modelled by setting up a program that manipulates them in a convenient way. In connectionist simulations, such variables emerge AFTER the simulation (otherwise, as we have seen, the analogy is lost). As a simple example, cluster analysis might reveal that different groups of units are activated under different input conditions (Greco & Cangelosi, 1996): analogy, in the form of structure-mapping, is in this case preserved [2]. All one is doing here is (a) considering what has happened inside a net that has been described, by analogy, as if it were performing a cognitive task and (b) interpreting it, by preserving the analogy "as if" it concerned cognitive variables.

18. Green would presumably reply that if this were taken seriously, such nets would no longer be connectionist but symbolic, like localist nets. Or that all their properties (nodes, connections, etc.) would just be another tool (however complex) for achieving symbolic representations. If Green's objective were still to attribute cognitive reality to nodes, activations, etc., then the above argument about the analogical use of models and the role of irrelevant details should be a satisfactory answer. But if this is not the case, if he is asking for something more, such as assigning to connectionist entities a deterministic causal role in the process (i.e., what I called a good analogy), one could still answer Green's challenge that there must exist "ontological realms that perhaps lie somewhere between the mental symbols of 'classical' cognitive science and neural activity" (Green 1998d, endnote 1). I will not invoke the well-known merits of connectionism, such as context sensitivity or graceful degradation, which make them plausible models of cognition. There are less appreciated reasons: in the way they work one can discover good analogies, impossible to see with traditional symbol systems.

19. Green's challenge is based on the (rather common) assumption that the only possible relationship between symbols and nonsymbols is implementation. In my view, this is a misconception deriving from the cognitivist metaphor of human information processing, so impressed in

our minds because of a strong tradition and the influence of authors such as Fodor & Pylyshyn (1988; Pylyshyn 1984). A full treatment of this question would be beyond the scope of this commentary, but I will try to sketch a possible alternative. The cognitivist tradition stems, in turn, from a rationalist philosophical tradition that would be called a "top-down" approach today. It poses the question, starting from language and all the complex symbols people use: what neural processes implement these? The computational metaphor fits well with this line of inquiry because in computation high-level variables can be seen naturally as "implemented" in low-level (physical) processing. A one-to-one correspondence is taken for granted. Each symbol should have a place in the "language of thought" (Fodor 1975), and since no one can deny that at bottom it must all be based on neural processes, there must be neural processes that "implement" such symbols. The wrong step in this line of reasoning was to assume that symbols are the starting point, something that exists from the outset and that only needs to be put in correspondence with some neural process.

20. A different line of reasoning becomes apparent when one raises the question: Where do symbols come from? Why should cognitive processes consist only of symbol manipulation? What about symbol ORIGINS? The idea that cognition somehow developed over time was originally explored by the Gestalt psychologists and has recently given rise to nonsymbolic

approaches to representation (Hatfield, 1989; see Greco, 1990a, 1990b, for discussion and other references). Neural nets may make it possible not to only implement but to originate symbols. This form of simulation was not possible with traditional AI symbol systems. We now have nets that behave "as if" symbolic activities were produced from nonsymbolic ones: Such models provide a substantive analogy with real cognitive processes that have not proven simulable by other means.

21. Let me conclude with one more respect in which connectionist nets could mark a turning point in cognitive modelling, concerning not only the content (whether analogies are clear and good) but the simulation method as a whole. In traditional symbolic models, symbols were interpreted only by the model builders and users, not by the models themselves. This is an important difference from natural cognition, where symbols alone, in a closed system, cannot have meaning unless they are grounded, i.e. connected to the world via nonsymbolic or presymbolic sensory information (Greco et al., 1998; Harnad 1990). This difference could turn out to be crucial, elucidating many problems arising in traditional modelling.

22. For example, symbolic models have proved to have severe shortcomings when one tried to move from "toy" tasks towards the natural complexity of the human mind. "Scaling up," it had been

thought, merely amounted to giving a system ever more abstract symbolic descriptions to allow it deal with ever more data. Explaining complex processes would just be a matter of reducing them to more elementary symbols to manipulate. One more aspect of the scaling problem is the so-called "frame problem" (McCarthy & Hayes, 1969; Ford & Hayes 1992; Harnad 1993), which arises in trying to give a formal account of what things do and do not change in dynamic situations. This difficulty really has to do with that of foreseeing and formally coding all the things that could happen to a system in the world (all the contingencies it could encounter); these all amount to potential data to the system. A system cannot in general deal with new symbols when they are not grounded in the system, but only in the programmer's head. So the real problem again concerns how to give a system the capability of constructing symbols through its own autonomous interactions with the environment, rather than trying to anticipate it all in advance in programmer-provided symbolic code.

23. In this commentary, I have tried to show that connectionist nets, as cognitive models, should be judged on the grounds of how clear the analogies are between what they do and what our cognitive system does. I have mentioned shortcomings that arise when this requirement is not fulfilled from the beginning, but I also pointed out benefits that come when analogies are established after interpretation. Moreover, I

suggested that there are even phenomena that may be essential to any cognitive simulation, such as symbol origins and symbol grounding, that can so far only be simulated using nets.

ENDNOTES

[1] The terms "cognitive" and "cognition" are used reluctantly, because I think this discussion should be addressed, more generally, to "cognitive or psychological" models, not cognitive phenomena only. "Mental" would be the best term.

[2] A second transposition concerns the validation phase that always followed a successful simulation in traditional symbolic AI simulations, verifying whether the hypothetical process implemented in the program actually produced the expected outcome. This validation could only apply to inputs and outputs (the Turing test was often invoked). In connectionist simulation, input/output validation is the first step; the description of internal processes comes later.

[3] This confusion is probably based on an overgeneral and incorrect use of the term "model" (Greco, 1994).

REFERENCES

Fodor, J. A. (1975) The language of thought New York Thomas Y. Crowell

Fodor, J. & Pylyshyn, Z. (1988) Connectionism and cognitive architecture: A critical analysis. Cognition 28: 3 - 71

Ford, K. M. & Hayes, P. J. Reasoning agents in a dynamic world: The frame problem. PSYCOLOQUY 3(59)

ftp://ftp.princeton.edu/pub/harnad/Psycoloquy/1992.volume.3/psycoloquy.92.3.59.frame-problem.1.ford+hayes

Goldsmith, M. (1998) Connectionist modeling and theorizing: Who does the explaining and how? PSYCOLOQUY 9(18)

ftp://ftp.princeton.edu/pub/harnad/Psycoloquy/1998.volume.9/psycoloquy.98.9.18.connectionist-explanation.15.goldsmith

Greco, A. (1990a). Some remarks about connectionism in psychological simulation. Cognitive Systems, 2-4, 359-372.

http://cogprints.soton.ac.uk/abs/psyc/199804024

Greco, A. (1990b). What kind of psychological processes can be modelled by a connectionist system? Paper presented at the Workshop "Connectionism: bridge between mind and brain?," Center for Interdisciplinary Research ZIF, Univ. of Bielefeld, Germany.

http://www.lettere.unige.it/sif/strutture/1/home/greco/zif90/zif90 .ht
m

Greco, A. (1994) Integrating "different" models in cognitive
psychology. Cognitive Systems, 4-1, 21-32.
http://cogprints.soton.ac.uk/abs/psyc/199804025

Greco, A., Cangelosi, A. (1996). Language, categorization, and
representation: a pilot study using neural networks. DISA Technical
Report n. 96-03, Univ. of Genova, Italy.
http://www.lettere.unige.it/sif/strutture/1/home/greco/techrep3.zi p

Greco, A., Cangelosi, A., Harnad, S. (1998) A connectionist model for
categorical perception and symbol grounding. Paper accepted at the
International Conference on Artificial Neural Networks (ICANN 98).
http://cogprints.soton.ac.uk/abs/psyc/199803023

Green, C. D. (1998a). Are connectionist models theories of cognition?
PSYCOLOQUY 9(4)
ftp://ftp.princeton.edu/pub/harnad/Psycoloquy/1998.volume.9/psyc.98.9.0
4.connectionist-explanation.1.green.

Green, C.D. (1998b) Connectionist nets are only good models if we know

what they model. Reply to Lee et al. on Connectionist- Explanation.

PSYCOLOQUY 9(23)

ftp://ftp.princeton.edu/pub/harnad/Psycoloquy/1998.volume.9/psycoloquy.98.9.23.connectionist-explanation.20.green

Green, C.D. (1998c) Higher functional properties do not solve

connectionism's problems. Reply to Goldsmith on Connectionist-

Explanation. PSYCOLOQUY 9(25)

ftp://ftp.princeton.edu/pub/harnad/Psycoloquy/1998.volume.9/psycoloquy.98.9.25.connectionist-explanation.22.green

Green, C.D. (1998d) The degrees of freedom would be tolerable if nodes

were neural. Reply to Lamm on Connectionist-Explanation. PSYCOLOQUY 9(26)

ftp://ftp.princeton.edu/pub/harnad/Psycoloquy/1998.volume.9/psycoloquy.98.9.26.connectionist-explanation.23.green

Harnad, S. (1990) The Symbol Grounding Problem. Physica D 42: 335-346.

http://cogprints.soton.ac.uk/abs/psyc/199803014

Harnad, S. (1993) Problems, Problems: The Frame Problem as a Symptom

of the Symbol Grounding Problem. PSYCOLOQUY 4(34)

ftp://ftp.princeton.edu/pub/harnad/Psycoloquy/1992.volume.3/psycoloquy.92.3.34.frame-problem.11.harnad

Hatfield, G. (1989). Computation, representation, and content in noncognitive theories of perception. In: Silvers, S. (Ed.), Rerepresentation. Dordrecht: Kluwer.

Hesse, M.B. (1966). Models and analogies in science. Notre Dame, Indiana: University of Notre Dame Press.

Lee, C., van Heuveln, B., Morrison, C.T., Dietrich, E. (1998). Why connectionist nets are good models. PSYCOLOQUY 9(17) ftp://ftp.princeton.edu/pub/harnad/Psycoloquy/1998.volume.9/psyc.98.9.17.connectionist-explanation.14.lee

McCarthy, J., Hayes, P. (1969) Some philosophical problems from the standpoint of artificial intelligence. In Meltzer B., Michie D. (eds.) Machine intelligence, vol. 4. New York: American Elsevier, pp.463-502.

Medler, D.A., Dawson, M.R.W. (1998). Connectionism and cognitive theories. Psycoloquy 9(11) ftp://ftp.princeton.edu/pub/harnad/Psycoloquy/1998.volume.9/psyc.98.9.11.connectionist-explanation.8.medler

Nagel, E. (1961).. The structure of science, Harcourt, Brace & World, New York.

Pylyshyn, Z. W. (1984) Computation and cognition. Cambridge MA: Bradford