mit decision as to whether a given theory of cognition captures them adequately. We just do not know that much as yet.

The authors say that the criteria deserve "greater scientific prominence." They therefore try to "evaluate theories by how well they do in meeting" the criteria (sect. 1, para. 4). This may be premature. Whether it is, depends on the merit of Newell's idea more than on its applications. So, it requires examination. What the authors call the Newell Test is a test not of Newell's idea but of the theories that should agree with it – provided that it is valid. Is it? How are we to judge this? (The claim that a characterization of "a functioning system of human cognition" is possible, is significant for those who plan to realize it. The claim and the plan pertain to studies of artificial intelligence that are no longer popular. See Agassi 1988, reprinted in Agassi 2003.)

Anderson & Lebiere (A&L) apply Newell's Test to two new ideas that are controversial, so the application cannot be such a test. Hence, their work is begging the question: Some test of it is required to show that it deserves a "greater scientific prominence."

"Newell is calling us to consider all the criteria and not pick and choose the ones to consider" (sect. 2.8). This remark renders the whole venture too questionable. The authors make it a propos discussion of the criterion of consciousness.

> Newell acknowledged the importance of consciousness to a full account of human cognition, although he felt compelled to remark that "it is not evident what functional role self-awareness plays in the total scheme of mind." We too have tended to regard consciousness as epiphenomenal . . . (sect. 2.8)

This is very shaky. Whether consciousness is or is not epiphenomenal is a red herring: It is an empirical fact that in many cases cognitive conduct differs depending on whether it is accompanied with consciousness or not, and the question may arise, should a system emulating human consciousness reflect this fact? Importantly, Turing's celebrated idea of the Turing Test is designed to avoid this question altogether.

The authors examine two sets, classical connectionism and ACT-R. Classical connectionism is a computerized version of behaviorism. (For the claim that connectionism is unscientific, see Agassi 1988, reprinted in Agassi 2003.) ACT-R is "a theory of cognition," "a subsymbolic activation-based memory" able "to interact with a symbolic system of production rules"; the "R" in ACT-R "denotes rational analysis" (sect. 4.1, first para.). The two sets, then, are artificial intelligence or expert-systems programs. The authors report a claim that classical connectionism passes the Turing Test. Presumably they disagree. The same holds for ACT-R. "ACT-R, except for its subsymbolic limitations, is Turing equivalent, as are most production systems" and "(proof for an early version of ACT appears in Anderson 1976)" (sect. 5.1, para. 2). This is a bit cryptic; I will explain the difference between the Turing and the Newell Tests in the following paragraph.

The Turing Test was meant to render the mind-body problem empirically decidable. Were there a computer program that could fool an expert, Turing suggested, then it would be empirically indistinguishable from humans, and so the attribution to humans of a metaphysical soul would be redundant. (For a critique of this test, see Agassi 1992.) Because Newell's criteria depict human characteristics, any interlocutor who can pass the Turing Test should certainly possess them, because the inability to exhibit any human characteristic the like of which Newell mentions would expose the impostor. And yet, the Turing Test is implicit and Newell's Test is explicit. This permits finding a partial success in passing the Newell Test. But, to be an explicit version of the Turing Test, the Newell Test must refer to a complete list of characteristics. We do not have this, and the Turing Test may be preferred just because it leaves this task to the experts who wish to test the humanity of their enigmatic interlocutor. Consequently, a Turing Test can never be decisive: Both expert and programmer can improve on prior situations and thus deem failure a merely temporary setback. True, the Turing Test is generally deemed possibly decisive, and, being a thought-experiment, actually decisive. Some writers, notably Daniel Dennett, claim that only exorbitant costs prevent the construction of a machine that will pass the Turing Test. That machine, then, should certainly be able to pass the Newell Test with flying colors. It is a pity that A&L do not refer to this claim and expose it as a sham. If they are anywhere close to being right, they should be able to do so with ease. (See Agassi & Laor 2000, pp. 556–62.)

The interesting aspect of the target article is that it is open-ended: Whenever the system A&L advocate, which is ACT-R, does not pass the examination as well as they wish, they recommend trying an improvement, leading to a retest. They should observe that such a move may be two-pronged. They refer to the improvement of the ability of a program to abide by the theory of flexibility; adaptive behavior; and the ability to integrate knowledge, use a natural language, and learn. They should not ignore the need to improve on these theories. When they refer to natural languages or to learning, they view the connectionist idea of them as more satisfactory than that of ACT-R, because it is more complete. Yet, whatever completeness is exactly, it is not enough: We seek explanations, and so to accept axiomatically what we want to understand is not good enough. We still do not know what a natural language is and how we learn; and we do not begin to understand these. Let me end with an insight of David Marr that should not be forgotten. Emulation is helpful for the understanding but is no substitute for it; sometimes, the very success of emulation, Marr observed, renders it less useful than a problematic one. We want understanding, not mere emulation. (See Marr 1982, p. 253 and Agassi 1987, p. 410.)

## Think globally, ask functionally

Erik M. Altmann

*Department of Psychology, Michigan State University, East Lansing, MI 48824.* **ema@msu.edu** **http://www.msu.edu/~ema**

**Abstract:** The notion of functionality is appropriately central to the Newell Test but is also critical at a lower level, in development of cognitive sub-theories. I illustrate, on one hand, how far this principle is from general acceptance among verbal theoreticians, and, on the other hand, how simulation models (here implemented within ACT-R) seem to drive the functional question automatically.

Anderson & Newell (A&L) have been carrying integrative cognitive theory, in shifts, for the past 30 years or so (if one goes back to Newell 1973). We are fortunate that Anderson is young; formulating dichotomous questions – seeing the trees but not the forest – may be the modal tenure procedure in psychology departments today, but perhaps in another generation it will be acceptable not to conduct new experiments at all but simply to integrate old data into increasingly complete computational models.

In the meantime, how can we avoid theoretical myopia in our daily research? Applying the Newell Test is well and good once a decade or so, with that many years' interim progress available to assess it. In terms of the next chunk of publishable research, however, it's useful to have more immediate guidance.

Central to the Newell Test is the idea of functionality: A theory has to explain how the cognitive system accomplishes some particular function. Among the Newell Test criteria, this function is high level, related in some relatively direct way to the fitness of the organism. However, as one develops micro-theories within a larger theory, functionality is still a relevant question; one can ask, for each process within a model, what it is for. Its outputs could, for example, be necessary inputs for another process in a chain that leads ultimately to accomplishing the task at hand. Or, one could ask whether each behavioral measure reflects a distinct process at all; perhaps it reflects a side effect of some other functionally necessary process. In both cases, it is difficult if not impossible to address the functional question without a precise representation of the processes one is talking about. In practice, this implies a computational simulation.



https://doi.org/10.1017/S0140525X03230132 Published online by Cambridge University Press

How does functionality play out at the level of the micro-theory that is the next chunk of publishable research? Curiously, even at this level, functionality seems to be regarded as optional, if not actually vulgar. A&L raise the example of short-term memory constructs (and Newell's frustration over them), but let's have a newer one, if only to see what might have changed. In the domain of executive control, there is a burgeoning literature on "switch cost" – the time cost associated with switching to a different task, as compared to performing the same task over again. One regularity to have emerged is that switch cost is difficult to erase; even with time and motivation to prepare for the other task, people are slower on the first trial under that task than on the second. The dominant theoretical account of this residual switch cost is arguably the "stimulus cued completion" hypothesis of Rogers and Monsell (1995, p. 224):

> This hypothesis proposes that an endogenous act of control deployed before onset of the stimulus can achieve only part of the process of task-set reconfiguration. Completion of the reconfiguration is triggered only by, and must wait upon, the presentation of a task-associated stimulus.

In terms of functionality, this hypothesis is vacuous. It need not be; one could ask how the system might benefit from stimulus-cued completion. For example, one could propose a benefit to the system hedging its bets and waiting to complete the reconfiguration process until there is evidence (in the form of the trial stimulus) that the new task set will be needed. One could then try to formulate scenarios in which this benefit would actually be realized and evaluate them for plausibility, or perhaps even against existing data. None of this was attempted by Rogers and Monsell, or by authors since who have invoked stimulus-cued completion as an explanatory construct. Call this a working definition of theoretical myopia: a "hypothesis" that merely relabels an empirical phenomenon.

In a subsequent ACT-R model, Sohn and Anderson (2001) explain residual switch cost in terms of stochasticity. Their model contains a "switching" production that retrieves the new task from memory and installs it in the system's focus of attention. Selection of productions is, like most other cognitive processes, subject to noise, which explains why this production is not always selected in advance of stimulus onset. Functionally, it can be selected after stimulus onset, though must be selected before response selection. This account is an improvement; it makes predictions (in terms of response-time variability), and it explains residual switch cost as a side-effect of the noise that accompanies any communication channel.

One could go further and ask, does residual switch cost reflect a process that directly contributes in some way to task performance? In another ACT-R model, Gray and I proposed that residual switch cost reflects a redundant task-encoding process that affects quality control (Altmann & Gray 2000). (Initial task encoding activates a memory trace for the current task, but noisily; redundant task encoding catches and properly strengthens memory traces that were initially weakly encoded.) The proof of functionality lay in Monte Carlo simulations showing that overall performance accuracy was higher with this redundant phase than without.

Are Sohn and Anderson right, or are Altmann and Gray? We have not found a behavioral test; perhaps neuroimaging will someday afford a diagnostic. I would predict, however, that the stimulus-cued completion hypothesis will not find its way into a precisely formulated cognitive theory, micro or otherwise, unless relevant functional questions are posed first.

ACKNOWLEDGMENT

This work was supported by an Office of Naval Research Grant N00014–03–1–0063.

## Nonclassical connectionism should enter the decathlon

Francisco Calvo Garzón

*Department of Philosophy, Indiana University, Bloomington, IN 47405, and University of Murcia, Facultad de Filosofía, Edif. Luis Vives, Campus de Espinardo, Murcia 30100, Spain.* **fjcalvo@um.es**

**Abstract:** In this commentary I explore nonclassical connectionism (NCC) as a coherent framework for evaluation in the spirit of the Newell Test. Focusing on knowledge integration, development, real-time performance, and flexible behavior, I argue that NCC's "within-theory rank ordering" would place subsymbolic modeling in a better position. Failure to adopt a symbolic level of thought cannot be interpreted as a weakness.

Granting Anderson & Lebiere's (A&L's) "cognitive decathlon" overall framework, and their proposed operationalizations and grading scheme for theory-evaluation, the aspects of their article that I address here concern the choice of contestants entering the decathlon, and, based on that choice, the exploration of nonclassical connectionism (NCC) as a coherent framework for evaluation in the spirit of the Newell Test. The range of classical connectionist architectures that A&L assess is confined to models that have a feedforward or a recurrent architecture, a locally supervised learning algorithm (e.g., backpropagation), and a simple nonlinear activation function (e.g., sigmoidal). A nonclassical framework, however, can be coherently developed. By NCC, I shall be referring to the class of models that have different combinations of pattern associator/autoassociative memory/competitive network topologies, with bidirectional connectivity and inhibitory competition, and that employ combined Hebbian and activation-phase learning algorithms (O'Reilly & Munakata 2000; Rolls & Treves 1998). Were NCC allowed to enter the competition, it would (or so I shall argue) obtain a "within-theory rank ordering" that could perhaps place it in a better position than the ACT-R theory. To demonstrate this, I will make three points with regard to 4 of the 12 functional constraints on the architecture of cognition that A&L take into consideration: knowledge integration, development, real-time performance, and flexible behavior.

On knowledge integration, classical connectionism (CC) gets a "worse" grade (see Table 1 of the target article). As an "intellectual combination" example of knowledge integration, A&L consider the literature on transfer of learning in infants. Marcus (2001) assessed the relationship between CC and rule-governed behavior by challenging the connectionist to account for experimental data that had been interpreted as showing that infants exploit (rule-governed) abstract knowledge in order to induce the implicit grammar common to different sequences of syllables (Marcus et al. 1999). Calvo and Colunga (in preparation) show how Marcus's infants-data challenge can be met with NCC (see Calvo & Colunga 2003, for a CC replica of this simulation). Our model (Fig. 1) is based on a simple recurrent network (SRN) architecture that has been supplemented with the following nonclassical features: (1) bidirectional (symmetric) propagation of activation, (2) inhibitory competition, (3) an error-driven form of learning (GenRec in McClelland 1994), and (4) the Hebbian model learning.

The fundamental component of our simulation resides in the fact that the network is pretrained with syllables that can be either duplicated or not. These first-order correlations in the environment amount to subregularities that can be exploited by the network in a semideterministic prediction task. During pretraining, the network learns to represent something general about duplication (i.e., sameness). This abstraction is crucial in encoding the patterns during the habituation phase. Like the infants in Marcus et al.'s study, the networks that were pretrained in a corpus in which some syllables were consistently duplicated learned to distinguish ABB patterns from ABA patterns after a brief period of training akin to infant's habituation.

Error-driven learning makes use of an activation-phase algorithm that, via bidirectional connectivity and symmetric weight



https://doi.org/10.1017/S0140525X03230132 Published online by Cambridge University Press