volved in the construction of attributions, whereas the amygdala and basal ganglia are responsible for trying to predict possible punishments and rewards related to one's actions (Lieberman et al. 2002; cf. Rolls 2000).

This system consists of a set of neural mechanisms tuned by a person's past experience and current goals; it is a subsymbolic, pattern-matching system that employs parallel distributed processing. It produces that continuous stream of consciousness we experience as "the world out there," whereas the rational system reacts to the spontaneous system, producing conscious thoughts experienced as reflections on the stream of consciousness (Lieberman et al. 2002). As a pattern-recognition system, the spontaneous system tries to combine all perceived features into a coherent representation; this is because the relevant neurons have been so paired by past experience that the activation of some will also activate others. The spontaneous system cannot consider the causal or conditional relationships between percepts because it does not operate by symbolic logic and because its links are bidirectional. Thus, simply asking a dispositional question (e.g., "Is this man prone to violent behavior?") may easily lead to an affirmative answer (Lieberman et al. 2002).

The rational system involves such brain areas as the anterior cingulate, prefrontal cortex, and hippocampus (Lieberman et al. 2002). It is a rule-based system able to encode any information that has a well-specified formal structure. Such a structure also allows the generation of new propositions on the basis of systematic inferences carried out in a language of thought which has a combinatorial syntax and semantics. It explicitly follows rules. This system thus seeks for logical, hierarchical, and causal-mechanical structure in its environment; operates on symbol manipulation; and derives knowledge from language, culture, and formal systems. It employs concrete, generic, and abstract concepts; abstracted features; compositional symbols; as well as causal, logical, and hierarchical relations. It is productive and systematic; abstracts relevant features; is strategic, not automatic; and serves such cognitive functions as deliberation, explanation, formal analysis, verification, ascription of purpose, and strategic memory (Sloman 1996).

The rational system either generates solutions to problems encountered by the spontaneous system, or it biases its processing in a variety of ways. A pre-existing doubt concerning the veracity of one's own inferences seems to be necessary for the activation of the rational system. The rational system thus identifies problems arising in the spontaneous system, takes control away from it, and remembers situations in which such control was previously required. These operations consist of generating and maintaining symbols in working memory, combining these symbols with rule-based logical schemes, and biasing the spontaneous system and motor systems to behave accordingly (Lieberman et al. 2002).

It could thus be argued that the spontaneous system is a collection of evolved mechanisms with an adaptive background, whereas computational universality is based on the ability of the rational system to exploit the evolved mechanisms to create algorithms for the performance of any cognitive task (see Pinker 1997, pp. 358–59; Atran 2002). This explains the fact that in many areas of everyday life people rely both on evolutionary intuitions and explicit theories. This distinction has recently been studied with regard to peoples' religious intuitions and their theological theories (e.g., Barrett 1998; 1999; Barrett & Keil 1996; Boyer 2001; Pyysiäinen 2003; Whitehouse 2002). Interaction between work on these types of real-life problem fields and on construction of hybrid systems might help us develop more integrated theories of human cognition.

ACKNOWLEDGMENTS

I thank Matthew Lieberman, Marjaana Lindeman, and Markku Niemivirta for help in writing this commentary.

The hardest test for a theory of cognition: The input test

Asim Roy

School of Information Systems, Arizona State University, Tempe, AZ 85287-3606. **asim.roy@asu.edu**

Abstract: This commentary defines an additional characteristic of human learning. The nature of this test is different from the ones by Newell: This is a hard, pass/fail type of test. Thus a theory of cognition cannot partially satisfy this test; it either conforms to the requirement fully, or it doesn't. If a theory of cognition cannot satisfy this property of human learning, then the theory is not valid at all.

The target article by Anderson & Lebiere (A&L) is very refreshing in the sense that it turns the focus back on accountability and tests for any theory of cognition. In examining theories of cognition, a look at system identification in science and engineering may be in order. In system identification, the basic idea is to construct an equivalent system (model) that can produce "behavior" that is similar to the actual system. So the key idea is to produce "matching external behavior." The equivalent system may not necessarily match the internal details of the system to be identified, but that is fine as long as it matches the system's external properties. And the external properties to match may be many. This is not to say that one should not take advantage of any information about the internals of the system.

Therefore, the crucial task for this science is to define the external behavioral characteristics that any system of cognition is supposed to exhibit. Understanding and characterizing the phenomenon to be modeled and explained is clearly the first and main step towards developing a theory for it. If that is not done, it is very likely that wrong theories will be proposed, because it is not known exactly what the theory should account for. This commentary defines an additional characteristic of human learning other than the ones in the Newell Test (Roy et al. 1997). In the spirit of the Newell Test, this is a characteristic of the brain that is "independent of" (1) any conjectures about the "internal" mechanisms of the brain (theories of cognition) and (2) the specific learning task. That is, this property of human learning is independent of a specific learning task like learning a language, mathematics, or a motor skill. The nature of this test is quite different from the ones provided by Newell: This is a hard, pass/fail type of test. In that sense, a theory of cognition cannot partially satisfy this test; it either conforms to its requirement fully, or it doesn't. This pass/fail test would allow one to quickly check the validity of alternative theories of cognition. If a theory of cognition cannot satisfy this property of human learning, then the theory is not valid at all. So this particular test is good enough for initial screening of theories. As explained in the following paragraphs, classical connectionism fails this test. One has to take a closer look at ACT-R and ACT-RN to pass judgment on them.

So what is this real hard test for theories of cognition? It can be summarized as follows: A brain-like system, constructed on the basis of some theory of cognition, is not permitted to use any inputs in its construction phase that are not normally supplied to a human brain. So the real hard test for any theory is in the inputs required to construct the relevant system of cognition. Let this test be called the "Input Test." The human brain has two sources of inputs during its development, both inside the womb and outside. Biological parents are the first source, and certain structures and systems can be inherited through that source. The other source of inputs for its development is the environment after birth. So any theory of cognition has to clearly delineate what pieces of its functioning system are inherited from biological parents and what pieces are developed subsequently through interactions with the environment. For humans, it is known for a fact that certain functionality of the brain is definitely not inherited, like the ability to speak a certain language, do mathematics, and so on. The modules for these functionalities/tasks do not come pre-built in the human brain; rather, they are developed and constructed gradually over time. So, to reiterate this point, the first task of a theory of cognition is to clearly delineate what pieces of its functioning system are inherited and what pieces are developed subsequently through interactions with the environment. And with regard to what can come pre-built (inherited), it has to provide sensible arguments.

Once a proposed theory of cognition maps out what is pre-built in the system in the sense of being inherited from biological parents, then the problem for the theory is to show how it develops and constructs the modules that are not pre-built. And whatever the means are for developing and constructing these modules, the hardest test for the theory is this: It has to demonstrate that it is not using any inputs for developing and constructing these modules that are not provided to humans from the environment. This input test can be explained nicely by examining classical connectionism. In classical connectionism, for example, network designs and other algorithmic information have to be externally supplied to the learning system, whereas no such information is ever an external input to the human brain. The well-known back-propagation algorithm of Rumelhart et al. (1986) is a case in point. In fact, many different network designs and other parameter values often have to be supplied to these learning systems on a trial-and-error basis in order for them to learn. However, as far as is known, no one has ever been able to externally supply any network designs or learning parameters to a human brain. So classical connectionism clearly violates this input test and is not a valid theory of cognition.

In general, for previously unknown tasks, the networks could not feasibly come predesigned in human brains; thus network designs cannot be inherited for every possible unknown learning problem faced by the brain on a regular basis. And the networks required for different tasks are different; it is not a one-size-fits-all situation. Since no information about the design of a network is ever supplied to the brain externally, it is therefore implied that the brain performs network designs internally. Thus, it is expected that any theory of cognition must also demonstrate the same ability to design networks and adjust its own learning parameters without any outside intervention. But the connectionist learning systems can't demonstrate this capability, and that again implies that classical connectionism is not a valid theory of cognition.

In summary, in this input test, a theory of cognition should be restricted to accepting information that is normally supplied to a human from the environment, nothing more.

Rethinking learning and development in the Newell Test

Sylvain Sirois

Department of Psychology, The University of Manchester, Manchester M13 9PL, United Kingdom. **sylvain.sirois@man.ac.uk**
http://www.psy.man.ac.uk/staff/sirois.htm

Abstract: The Newell Test is an ambitious and promising project, but not without pitfalls. Some of the current criteria are not theoretically neutral, whereas others are unhelpful. To improve the test, the learning and development criteria are reviewed and revised, which suggests adding a maturation criterion as well. Such changes should make the Newell Test more general and useful.

Anderson & Lebiere (A&L) have certainly embarked on an ambitious project: to transform Newell's (1980; 1990) functional criteria for human cognitive architectures into the ultimate test of cognitive theories. I certainly sympathise with such ambitions, especially given their emphasis on the functional aspects of the criteria that should be used. For example, we recently conducted a similar (albeit substantially more humble) exercise for models of infant habituation (Sirois & Mareschal 2002). We identified a set of seven behavioural and neural criteria that functional models of the phenomena need to satisfy. This proved extremely useful to highlight the limitations of current models, but also (and perhaps more importantly) to suggest what the next generation of models needed to address. Given the relatively small scale of the problem addressed in our work, one could conceivably expect huge and varied rewards from A&L's far more valiant endeavour.

Whereas the rewards may prove an exponential function of those we observe in analogous but restricted projects, so may the problems. The authors quite rightly acknowledge that their criteria (which are a slightly modified version of Newell's) are not the only criteria by which a theory can be assessed. But far more crucial than how many criteria (which makes the test more or less liberal) is the question of *which* criteria (which makes the test more or less useful). If the stated goal of such a test is to avoid theoretical myopia, then a few of the criteria are certainly problematic because they either imply that a model adheres to a specific school of thought or to tests of models against highly disputable standards. For example, *knowledge integration* may have been retitled from Newell (1990) but owes no less to symbolic tradition than when it was proposed by Newell. As such, the grading of this criterion is unduly biased towards models and theories originating from this tradition. The *consciousness* criterion is even more problematic: Whether the criterion has any functional value depends on an eventual theory that would make such a suggestion!

Other commentators will likely address the relevance or appropriateness of the various criteria, if not of the test itself. Despite inherent difficulties in such projects, I believe that a revised formulation of the Newell Test could be quite useful. I would thus like to focus on two criteria that, in my view, should be kept in the Newell Test: *learning* and *development.* Surprisingly, the authors evacuated the functional role of learning in their discussion. Moreover, they discuss development as a (perhaps functional) constraint rather than as a functional mechanism. In fact, what they present as development sounds remarkably like maturation.

The authors should not be blamed too harshly for reproducing a common problem in developmental psychology: confounding learning and development by discussing them in terms of *outcomes* rather than *mechanisms* (Liben 1987). This is most explicit when they present the slow learning of *classical connectionism* as satisfying the development criterion. If anything, and contrary to what the authors suggested, the sort of learning in classical connectionism can be characterised as a nativist learning theory (Quartz 1993; Sirois & Shultz 1999).

Fortunately, the notions of learning and development can be expressed formally as non-overlapping functions (Sirois & Shultz 1999). *Learning* can be defined as parametric changes that enable a given processing structure to adapt to its environment. *Development,* however, can be defined as structural changes that foster more complex adaptations, given learning failure. These definitions not only constrain the contribution of each mechanism to cognitive change, but also specify the relationship between learning and development. Learning causes the current structure to adapt, but when that fails, development alters the structure to promote further learning. It must be noted that either form of change is a function of experience. Within this framework, then, *maturation* becomes an experience-independent structural change that delays learning, in line with what A&L discussed as development.

Like others (including A&L), I believe that an adequate theoretical formulation of cognition must be consistent with learning and developmental issues. Moreover, given the significant changes that can be introduced by maturation (i.e., the cognitive structure increases in complexity), I would suggest that the Newell Test also incorporates maturation as one of its criteria. The grading is relatively straightforward for the learning, development, and maturation criteria. If a theory allows for parametric changes as a function of experience, it can learn. If it allows for experience-dependent structural changes that support further learning, it satisfies development. Finally, if it allows for experience-independent, programmed structural changes that modify the learning space, it satisfies maturation.

https://doi.org/10.1017/S0140525X03350137 Published online by Cambridge University Press