man 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.

These learning, development, and maturation criteria are general by design, and so are the grading proposals, in line with Newell's wish to avoid theoretical myopia. A cognitive theory should be granted with the ability to satisfy any of these criteria if it satisfies the relevant functional properties, irrespective of how the mechanisms are actually realised. This general nature does not imply that the criteria are vague, however. We initially proposed these definitions to discuss various classes of neural networks as they are applied to developmental problems. We found that the classical connectionist framework only satisfied the learning criteria (Sirois & Shultz 1999). But we applied the same framework to discuss the various mechanisms of Piagetian theory, clarifying them in the process, and allowing for a formal distinction between learning and developmental notions in Piaget's work (Sirois & Shultz 2003). If we apply these definitions to ACT-R as discussed by A&L, we could grant ACT-R the ability to satisfy learning and developmental criteria (the latter through the construction of new rules).

To summarise, the idea of a Newell Test is quite attractive but not without design pitfalls. Whereas there may be some inadvertent myopia in the choice of criteria, most of these may well be retained (but perhaps reformulated). The peer commentaries in this journal will hopefully provide the next few steps towards the design of a generally satisfying test of cognitive theories.

ACKNOWLEDGMENT
I thank Isabelle Blanchette for useful comments on an earlier draft.

What about embodiment?

David Spurrett
Philosophy Department, University of Natal, Durban, 4041, South Africa.
spurrett@nu.ac.za http://www.nu.ac.za/undphil/spurrett/

Abstract: I present reasons for adding an *embodiment* criterion to the list defended by Anderson & Lebiere (A&L). I also entertain a likely objection contending that embodiment is merely a type of *dynamic behavior* and is therefore covered by the target article. In either case, it turns out that neither connectionism nor ACT-R do particularly well when it comes to embodiment.

The principle that cognitive theories should be evaluated according to multiple criteria is worth adopting, and Anderson & Lebiere's (A&L's) development of Newell's proposals in this regard is useful. One important criterion seems to be missing, though, and that is *embodiment.*

By embodiment, I understand, loosely, physical implementation in an environment. Humans, clearly a key consideration of the target article, are, of course, embodied. They exhibit striking virtuosity at moving around the world and exploiting the resources available in it. Perhaps more important for present purposes, we are talented at exploiting the structure of environments (and of our bodies in them) for cognitive ends, or as some would have it, engaging in "distributed cognition" (e.g., Hutchins 1995). One example is locomotion, where recent research (Thelen & Smith 1994) indicates that the architecture of the body, and the properties of the body in interaction with the environment, play significant roles in control of behavior. Another example, rather closer to the concerns of traditional cognitive science, is the game of Tetris, where it has been shown (Kirsh & Maglio 1994) that human players use external actions to improve the efficiency (speed, accuracy, error rate) of the spatial manipulations and judgements demanded by the game. External rotation of a Tetris piece, along with inspection to establish whether the rotated piece is in a preferable orientation (compared to before), is often faster and less error-prone than mental rotation for the same purpose. This suggests that at least some cognitive problems are tackled using a coalition of internal and external resources, and that an important feature of our cognitive makeup is that we can detect opportunities for this. (Further examples in humans, other animals, and (some) robots abound. Clark [1997] is a useful survey.) This in turn indicates that a theory of cognition that fails to take embodiment seriously is unlikely to capture such features of our own cognitive performance.

A likely objection here notes that A&L's criterion 5 is "dynamic behavior." Since this criterion concerns the relationship between a cognitive system and an environment, perhaps, properly understood, it includes embodiment and distributed cognition. Distributed cognition just *is,* the objection goes – a kind of dynamical coupling between an information-processing system and a structured body and environment. This objection may be taking charitable interpretation too far. A&L's discussion of their "dynamic behavior" criterion (sect. 2.5 of the target article) places considerable emphasis on dealing with the unexpected, and relatively less on exploiting external structure. When evaluating the relative performance of classical connectionism and ACT-R with respect to the dynamic behavior criterion (sect. 5.5 of the target article), their emphasis is on real-time control, not embodiment. Rather than try to settle the question whether embodiment is or is not a version of dynamic behavior, I propose to consider how connectionism and ACT-R fare in the case where embodiment is added as a separate criterion, and where dynamic behavior is interpreted to include it.

Were embodiment added as a criterion, I suggest that connectionism would achieve mixed results. In some cases it does extraordinarily well. Consider Quinn and Espenschied's (1993) neural network for controlling a hexapod robot. The success of this system depends to a significant extent on allowing features of the physical construction of the robot, in interaction with the environment, to play a role in control – so that the motion of individual feet will be inhibited if other specific feet do not yet have secure positions. One way of understanding this is to regard the changing physical links between some neurons, parts of the robot anatomy, the physical environment, other parts of the anatomy and (eventually, and sometimes) other neurons, as functioning like additional neurons, or interneuron connections, transforming or transmitting information about footing, load on joints, and so on. In other cases, though, it is not (yet) clear how to go about building a network, embodied or otherwise, to handle tasks (such as air traffic control) involving fairly specific and detailed functional decomposition, tasks for which systems such as ACT-R seem well suited.

ACT-R, I argue, scores worse for embodiment. Its successes at, for example, modelling driving are in constrained simulation environments, where embodied interaction with the "feel" of the vehicle and its relation to the road surface, are absent, and where attendant opportunities for exploiting environmental structure (engine tone, vibration) to help cue such actions as gear changes are absent for both the human subjects who provide the target data, and the ACT-R models of driving behavior which do well at approximating the behavior of such humans.

However, we might reinterpret A&L's "dynamical behavior" criterion in a way that includes embodiment as a subtype of dynamic behavior. In this case, and in the light of what is said in the target article and so far in this commentary, connectionism should retain its mixed score. In this case ACT-R should also, I argue, receive a mixed score: It doesn't do well at plain embodiment, but does better at non-embodied forms of dynamic behavior. In either case, the moral to draw is that if embodiment is a genuinely important criterion, then *neither* connectionism nor ACT-R seem, as they stand, in a good position to perform consistently well on it.

https://doi.org/10.1017/S0140525X03360133 Published online by Cambridge University Press