rather than an established theory. It allows one to construct a wide variety of models and theories, but their efficiency should be tested against various criteria in order to construct adequate models. To some extent, this principle corresponds to the idea that various physical phenomena stem from the interactions among atoms or molecules. Before 1905, when Einstein proved the existence of these particles, this idea was also a heuristic, but its usefulness for physics is obvious.

However, the idea itself is not sufficient to construct physical models, so these interactions must correspond to various physical laws, such as the laws of thermodynamics. In a similar vein, the self-organizing models of cognition initially must be tested against some criteria relating to the properties of its architecture. Such micro-criteria seem absent (or not stated explicitly) in the target article; however, without using them, the comparison against macro-criteria such as Newell's is hardly correct because of the considerable arbitrariness in the models constructed. For instance, different models can merely describe various levels of the phenomenon under study.

Of course, the theory of cognition still does not have such strict laws as in physics, but several micro-criteria appear useful to judge self-organizing models. The first micro-criterion is the similarity in relevant brain functioning. Since self-organizing models of cognition implicitly refer to self-organizing brain activity which can involve various levels of brain functioning, various models can be compared if their architecture meets the same levels of brain functioning. The architecture of connectionism meets the level of single neurons, but the ACT-R architecture corresponds to cortical regions.

The second micro-criterion is the similarity in the determination of initial settings. Various models can be compared when similar efforts are necessary to establish their initial settings and these settings are equally robust to their changes. The robustness of connectionist settings is well known; ACT-R seems to require more precise but vulnerable settings. For example, the ACT-R model of learning the past tense in English (Taatgen & Anderson 2002) performs well, but the model seems to be vulnerable to the choice of the production rules and learning mechanisms used. It is not obvious that the model with slightly different characteristics could show similar results.

The last micro-criterion assumes that the complexity of entities, interactions, and conditions must be approximately the same in the models judged, or the architecture of one model must naturally result from emergent processes in the other. The architecture of connectionist models is simpler than ACT-R's and, realizing this, A&L describe another model, ACT-RN, which implements ACT-R by standard connectionist methods. However, this implementation seems artificial, for A&L simply predetermine the existence of ACT-R's slots and production rules instead of deriving them from more primitive features of a connectionist model. In principle, A&L simply demonstrate that ACT-RN (and, accordingly, ACT-R) meets the principle of self-organization.

One can conclude that three micro-criteria separate connectionism from ACT-R; these theories describe different levels of cognition; therefore, their direct comparison is hardly correct.

Dual-process theories and hybrid systems

Ilkka Pyysiäinen

Helsinki Collegium for Advanced Studies, University of Helsinki, FIN-00014, Finland. **ilkka.pyysiainen@helsinki.fi**
http://www.helsinki.fi/collegium/eng/staff.htm

Abstract: The distinction between such differing approaches to cognition as connectionism and rule-based models is paralleled by a distinction between two basic modes of cognition postulated in the so-called dual-process theories. Integrating these theories with insights from hybrid systems might help solve the dilemma of combining the demands of evolutionary plausibility and computational universality. No single approach alone can achieve this.

Not only are cognitive scientific "paradigms" disconnected; it also seems to be difficult for a theory of cognition to meet both Newell's criteria 1 and 11. An evolved cognitive architecture apparently cannot be computationally universal (e.g., Bringsjord 2001). Anderson & Lebiere (A&L) thus emphasize that humans can learn to perform almost arbitrary cognitive tasks, but they do not explain why some tasks are easier to learn than others. They suggest that applying a broad enough range of criteria might help us construct an exhaustive theory of cognition, referring to Sun's (1994; 2002) hybrid systems integrating connectionism and a rule-based approach as an example (see also Sun & Bookman 1995). I argue that the distinction between connectionist and functionalist models is paralleled by a distinction between two types of actual cognitive processing, as postulated within the so-called dual-process theories. These theories, developed in social psychology, personality psychology, and neuropsychology, for example, strongly suggest that there are two different ways of processing information, variously labeled

Intuition and implicit learning versus deliberative, analytic strategy (Lieberman 2000);

A reflexive and a reflective system (Lieberman et al. 2002);

Associative versus rule-based systems (Sloman 1996; 1999);

An experiential or intuitive versus a rational mode of thinking (Denes-Raj & Epstein 1994; Epstein & Pacini 1999; Epstein et al. 1992; Simon et al. 1997);

An effortless processing mode that works through associative retrieval or pattern completion in the slow-learning system elicited by a salient cue versus a more laborious processing mode that involves the intentional retrieval of explicit, symbolically represented rules from either of the two memory systems to guide processing (Smith & DeCoster 2000);

Implicit versus explicit cognition (Holyoak & Spellman 1993);

Intuitive versus reflective beliefs (Cosmides & Tooby 2000a; Sperber 1997).

Although the terminologies vary, there is considerable overlap in the substance of these distinctions. The two systems serve different functions and are applied to differing problem domains. They also have different rules of operation, correlate with different kinds of experiences, and are carried out by different brain systems. Some consider these two mechanisms as endpoints on a continuum, whereas Lieberman et al. (2002) argue that they are autonomous systems (see, e.g., Chaiken & Trope 1999; Holyoak & Spellman 1993).

By synthesizing the extant theories, with a special focus on Sloman (1996) and Lieberman et al. (2002), we may characterize the spontaneous system as follows. It operates reflexively, draws inferences, and makes predictions on the basis of temporal relations and similarity; and employs knowledge derived from personal experience, concrete and generic concepts, images, stereotypes, feature sets, associative relations, similarity-based generalization, and automatic processing. It serves such cognitive functions as intuition, fantasy, creativity, imagination, visual recognition, and associative memory (see especially, Sloman 1996). It involves such brain areas as the lateral temporal cortex, amygdala, and basal ganglia. The lateral temporal cortex is, for example, most directly in-

https://doi.org/10.1017/S0140525X03340130 Published online by Cambridge University Press

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 hu-

https://doi.org/10.1017/S0140525X03340130 Published online by Cambridge University Press