able to model or causally explain observable behaviour. Response times, button presses, verbal reports, and the like, must be the basis of any such theory; without such third-person information, a cognitive science theory would have nothing to explain.

Returning to the problem of consciousness (or the mind-body problem): Why do certain cognitive and emotional processes have specific experiential or so-called qualitative features? Block (1995) has argued for a difference between so-called access-consciousness (A) and phenomenal consciousness (P). A mental state is A-conscious if it can be poised as premise in reasoning, rational control of action and speech. A mental state is P-conscious if there is something it is *like* to be in that state (Nagel 1974). The mind-body problem is, then, normally interpreted as a problem of explaining how P is related to (other) physical matter.

Any cognitive theory should be able to explain or model what happens when subjects report about consciousness, or about anything else, for that matter. In themselves, however, such explanations or modelling exercises do not necessarily point at anything more than correlations between two sets of psychological third-person data, for example, verbal reports and brain activity. At best, this will give us an understanding of A-consciousness, but not necessarily of P. When describing a cognitive process in terms of its functions or causal processes, P does not fit in unproblematically. Even when turning to some of the more optimistic accounts, one finds arguments that cognitive science can *inform* a solving of the mind-body problem but not actually solve it (Overgaard 2003). Epistemologically speaking, one can easily describe one's experiences exactly without ever referring to the kinds of descriptions and models used by cognitive scientists. Vice versa, one can make a full description of a cognitive process in terms of mathematical models or the often-seen "boxes with arrows between them" without ever referring to experiential qualities. On this basis, one might reasonably question whether an explanation of consciousness is a realistic goal for cognitive science.

For this reason, we are sceptical of one basic supposition underlying the A&L target article: that the maximally broad "encompassing of its subject matter – the behavior of man" (Newell 1973, p. 288, cited in sect. 6, Conclusion, last para.) shall be regarded as an unquestioned quality criterion for theoretical models guiding cognitive research. On the contrary, one might argue that it would be a more theoretically sound approach to explicitly specify the limitations of a given paradigm and its possible openness and connectedness with other paradigms, rather than trying to extend it to apply to as many domains as possible.

The one existing type of language in which *everything* can be spoken about is natural, everyday language. The all-encompassing semantic capacity of natural, everyday language is bought at the price of a low degree of specificity as far as the identification of statements' truth conditions is concerned. The potential utility value of theoretical languages lies in their capacity to isolate and specify knowledge domains characterised by high degrees of epistemic consistency (for scientific purposes) and action predictability (for technological purposes). Definitely, at this stage of cognitive science, we fear this utility value may become jeopardised if success in theory building gets simplistically equated with breadth of coverage.

## Connectionism, ACT-R, and the principle of self-organization

Pavel N. Prudkov

*Ecomon Ltd., Selskohosyastvennaya str 12-A, Moscow, Russia.*
**Pnprudkov@mtu-net.ru**

**Abstract:** The target article is based upon the principle that complex mental phenomena result from the interactions among some elementary entities. Connectionist nodes and ACT-R's production rules can be considered as such entities. However, before testing against Newell's macro-criteria, self-organizing models must be tested against criteria relating to the properties of their elementary entities. When such micro-criteria are considered, they separate connectionism from ACT-R and the comparison of these theories against Newell's Tests is hardly correct.

The target article by Anderson & Lebiere (A&L) is devoted to the demonstration of the possibilities of the ACT-R theory. To this end, the authors compare their theory against connectionism on the basis of Newell's criteria for a theory of cognition. However, it is difficult to understand from the article why A&L have decided to select connectionism as a competitor of ACT-R. Indeed, if ACT-R is an unified framework, but the term "connectionism" is "used in the field to refer to a wide variety of often incompatible theoretical perspectives" (target article, sect. 3, para. 7), then A&L could test ACT-R against, for example, several symbolic models sharing certain common characteristics.

It seems that the main reason for A&L's choice (acknowledged only partially by A&L) is the principle of self-organization, that is, the assumption that complex mental phenomena can be described as a result of the interactions among some elementary entities. This principle has been suggested by me elsewhere (cf. Prudkov 1994), and it was based on the following two facts. First, we know that mental processes are heavily connected to various aspects of brain functioning, though the mechanism of this connection is still unclear. Second, neuroscience data demonstrate that the complex forms of brain activity result from the interactions among some elementary brain entities. Brain areas, single neurons, parts of a neuron, distributions of electrical fields, and the like, can be treated as such entities in accordance with the level of brain functioning considered. It seems impossible to reduce all neural levels to a basic one.

The principle of self-organization requires no correspondence between cognitive elementary entities and any of their neural counterparts, though such correspondence is possible. But all characteristics of a cognitive self-organizing process must result from the properties of its elementary entities and interactions among them, without involving any factors external to the system. The architecture of a self-organizing system is defined by three sorts of characteristics (Prudkov 1994). First, it is necessary to define the elementary entities of the system. Second, the results of the interactions between the entities must be determined. Because the idea of interaction supposes changes in components of the entities, one can say self-organizing models by definition are hybrid. And, third, all conditions or probabilities of the interactions to occur must be described. Learning, then, corresponds to long-term changes in a self-organizing system.

With connectionist nodes as elementary entities, it is intuitively clear that connectionism complies with the principle (a more detailed representation is in Prudkov 1994). With the biological implausibility of many connectionist methods, the principle is likely to be the main reason to use connectionism for understanding cognition (Green 1998). To convert the ACT-R theory into self-organization terms, suppose that production rules are elementary entities, matching the conditions of production rules, and the state of declarative memory determines which entities can interact at a given time. Finally, the rule selected for firing, the result of the firing along with the corresponding changes in declarative memory, is the consequence of an interaction.

Of course, this principle must be considered as a heuristic 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-