to the development of language and cognition, scant attention has been given to the issue of biological plausibility in discussions of innate properties, and there has been little investigation of the potential variety of ways in which something could be innate. In contrast, and as a direct result of their experience with connectionist models, Elman et al. (1996) not only present a case against the plausibility of "representational nativism," but also offer a framework for developing alternative conceptions of innate constraints on development that draws on architectural and timing constraints in connectionist models as a guide.

In addition to clarifying the necessary conditions for development, connectionist models also provide a vehicle for exploring the *dynamics* of development. One of the key insights provided by connectionist models is that the mapping between overt behaviour and underlying mechanism is often nonlinear. As Elman et al. (1996) emphasize, contrary to assumptions underpinning much developmental research, qualitative changes in behaviour do not necessarily signal qualitative changes in the mechanisms responsible for that behaviour. Instead, these models demonstrate that sudden dramatic effects in terms of the output of a system can be produced by tiny, incremental changes in internal processing over time. In the case of ontogenetic development, this suggests that apparent discontinuities in conceptual or linguistic understanding or output may not be the result of new mechanisms coming online at certain points in development as has often been assumed, but instead reflect the continuous operation of the same mechanism over time.

Added to demonstrations of how the same mechanism can be responsible for multiple behaviours, connectionist models can also illuminate the reverse case in which a single outcome or behaviour arises through the action of multiple interacting mechanisms. Further, Elman et al. (1996) point to instances where the same behavioural outcome can be produced in a number of different ways, as in the case of degraded performance in artificial neural networks. (See Karmiloff-Smith 1998 for how crucial this is in understanding so-called behaviour in the normal range in some developmental disorders.) Precisely because connectionist models allow researchers to probe the potential range of relations that can exist between behavioural outcomes and their underlying causes, they overturn assumptions of straightforward one-to-one mapping between mechanisms and behaviour and are therefore useful in revealing the "multiplicity underlying unity" in development (Elman et al. 1996, p. 363).

The preceding are but a few examples that identify specific issues in developmental psychology where connectionist tools have demonstrated natural applications. More generally, the resources of connectionism have also been a critical factor in recent attempts to develop a viable interactionist framework for cognitive developmental research. Commenting on the connectionist inspired framework advocated by Elman et al. (1996), Newcombe (1998) points to a recent trend in cognitive developmental theorising that eschews the extremes of nativist and empiricist approaches to learning and cognition, in favour of an account that offers some substantive ideas about the reciprocal actions of organism and environment in producing developmental change. From this standpoint, the resources of connectionism can be seen to contribute to this project by offering researchers a specified, formal account of the developmental process that goes well beyond the verbal accounts typical of developmental theory. Moreover, as Elman et al. (1996) point out, the striking resemblance between the process of error reduction in artificial neural networks and earlier attempts to depict epigenesis in natural systems (e.g., Waddington 1975) offers further evidence of the utility of connectionism for attempts to formalize the interactional nature of development.

The preceding sketch serves to highlight some of the variety of ways in which the computational and conceptual resources of connectionism have been usefully applied in developmental psychology. Yet these pragmatic benefits of connectionist models are not readily apparent in A&L's present evaluation of connectionism against the Newell Test designed to reveal an adequate theory of cognition. As it stands, their evaluation falls short of a comprehensive comparative appraisal of ACT-R as a candidate theory of cognition, and it fails to bring forth the utility of the connectionist toolbox for cognitive science research.

On the encompassing of the behaviour of man

Morten Overgaard[a] and Soeren Willert[b]
[a]*Department of Psychology, University of Aarhus, Asylvej 4, Risskov 8240, Denmark;* [b]*Center for System Development, University of Aarhus, Katrinebjergvej 89GAarhus N 8200, Denmark.* **Overgard@pet.au.dk**
swi@psy.au.dk **www.psy.au.dk/phd/morten**
www.psy.au.dk/ompi/soeren

Abstract: One supposition underlying the Anderson & Lebiere (A&L) target article is that the maximally broad "encompassing of its subject matter – the behavior of man" (cf. sect. 6, last para.) is regarded as an unquestioned quality criterion for guiding cognitive research. One might argue for an explicit specification of the limitations of a given paradigm, rather than extending it to apply to as many domains as possible.

Anderson & Lebiere (A&L) set out on an important and admirable mission: to evaluate theories within the more or less well-defined area of cognitive science from one set of criteria in order to avoid a dissolving of theories into disconnected paradigms. We shall not criticise their general idea of measuring comparable theories with a common yardstick, nor the actual grading of ACT-R and connectionism presented by A&L. However, the very approach implies that there is a set of theories that can legitimately be labelled "cognitive theories." To decide whether a given theory falls under the category "cognitive science" and thus decide which theories it would be meaningful to grade with the Newell Test, certain basic requirements must be fulfilled. One could ask whether such basic requirements would be identical to the criteria in the A&L version of the Newell Test. If that were indeed the case, we could have no theory that could truly be called *cognitive* to this day. For instance, we have no theory to explain why consciousness is "a functional aspect of cognition" (let alone one that also explains dynamic behaviour, knowledge integration, etc.) (Chalmers 1996; Velmans 1991). Furthermore, it would be a circular enterprise indeed to measure a theory according to criteria identical to the ones it must already fulfil.

Most likely, however, one would not equate the basic requirements for cognitive science with the criteria of the Newell Test. For such a purpose, the criteria seem to be set much too high. Rather, one would look at the many *different* usages of the term *cognitive* within the research field in general and establish relevant criteria on this basis. This, however, leads us into the situation where we presently stand, that is, a situation where "cognitive science" is loosely defined. We have a number of core theories that definitely are cognitive – such as Treisman's attenuation model (Treisman & Gelade 1980) or the SAS model of visual attention (Norman & Shallice 1986) – and several borderline cases – such as Gibson's ecological perception theory (Gibson 1979) – where it is unclear whether the theory is truly a cognitive psychological theory.

Although our conceptualisation of cognitive science does not seem very exact, it seems safe to say that it has developed historically as an attempt to explain the transition from stimulus to response by "internal variables" (see Tolman 1948). Thus, all cognitive theories – the core cases as well as the less clear-cut ones – intend to give explanations in terms of functions. No matter how the specific theories are construed, all cognitive theories explain the function of some mental phenomenon, whether they collect empirical data from behavioural measures, computer simulations, mathematical models, or brain scannings. This common point of departure has certain consequences for the kind of theory that can be developed. First and foremost, any cognitive theory must be 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