models have achieved a higher degree of fidelity to the actual properties of human cognition. What this indicates to us is that, although the shortcomings of symbolic models may be temporary (as A&L suppose), they are most likely to be overcome by incorporation of the very principles that govern processing as defined at the connectionist level.

Second, as symbolic modelers take each new step in the direction of connectionist models, they do so accepting the fact that the phenomena to be explained have the characteristics that served to motivate the exploration of connectionist models in the first place. This, in turn, undermines the stance that the fundamental principles of human cognition should be formulated at the symbolic level, and instead further motivates the exploration of principles at the connectionist level. While we acknowledge that connectionist models still have many limitations, we nevertheless feel that this does not arise from any failure to acknowledge a symbolic level of thought. Instead we suggest that it arises from the fact the connectionists (like symbolic modelers) have not yet had the chance to address all aspects of cognition or all factors that may affect it.

In spite of our feeling that the facts of human cognition are completely consistent with the principles of parallel distributed processing, we do not wish to give the impression that we see no merit in modeling that is directed at the symbolic level. Given that symbolic formulations often do provide fairly good approximations, it may be useful to employ them in cases where it would be helpful to exploit their greater degree of abstraction and succinctness. We believe that work at a symbolic level will proceed most effectively if it is understood to be approximating an underlying system that is much more parallel and distributed, because at that point insights from work at the connectionist level will flow even more freely into efforts to capture aspects of cognition at the symbolic level.

ACKNOWLEDGMENTS

Preparation of this commentary was supported by Interdisciplinary Behavioral Science Center Grant MH 64445 from the National Institute of Mental Health (USA). Tiago V. Maia was supported by the Foundation for Science and Technology (Portugal). We thank the other members of the PDP Research Group at Carnegie Mellon for useful discussions.

NOTES

1. It is necessary to note that none of the models we have discussed fully embody all the principles of the PDP framework. For example, the interactive activation and TRACE models use localist, not distributed, representations, while the models of spelling-to-sound mapping (Seidenberg & McClelland 1989; Plaut et al. 1996) do not incorporate intrinsic variability. This fact can lead to confusion about whether there is indeed a theoretical commitment to a common set of principles.

In fact, we do have such a commitment. The fact that individual models do not conform to all of the principles is a matter of simplification. This leads to computational tractability and can foster understanding, and we adopt the practices only for these reasons. Everyone should be aware that models that are simplified embodiments of the theory do not demonstrate that models incorporating all of its complexity will be successful. In such cases further research is necessary, especially when the possibility of success is controversial. For example, Joanisse and Seidenberg (1999) used localist word units in their model of past-tense inflection, and Pinker and Ullman (2002a; 2002b) have argued that this is essential. In this context, we fully accept that further work is necessary to demonstrate that a model using distributed semantic representations can actually account for the data.

2. It should be noted here that none of these models assume that learning occurs through correction of overtly generated errors. Instead, it is assumed that exposure provides examples of appropriate usage in context. The learner uses the context as input to generate an internal representation corresponding to the expected phonological form. Learning is driven by the discrepancy between this internal representation and the actual perceived form provided by the example.

3. Marcus et al. (1995) claimed that German has a regular plural (the so-called +s plural) that conforms to the expectation of the symbolic approach, in spite of the fact that it is relatively infrequent. However, subsequent investigations indicate that the +s plural does not exhibit the properties one would expect if it were based on a symbolic rule (Bybee 1995; Hahn & Nakisa 2000).

Evaluating connectionism: A developmental perspective

Claire F. O'Loughlin[a] and Annette Karmiloff-Smith[b]

[a]*Department of Psychology, University of Canterbury, Christchurch, New Zealand;* [b]*Neurocognitive Development Unit, Institute of Child Health, London WC1N 1EH, UK.* **aallardy@xtra.co.nz a.karmiloff-smith@ich.ucl.ac.uk**

Abstract: This commentary questions the applicability of the Newell Test for evaluating the utility of connectionism. Rather than being a specific theory of cognition (because connectionism can be used to model nativist, behaviorist, or constructivist theories), connectionism, we argue, offers researchers a collection of computational and conceptual tools that are particularly useful for investigating and rendering specific fundamental issues of human development. These benefits of connectionism are not well captured by evaluating it against Newell's criteria for a unified theory of cognition.

In this commentary, we question Anderson & Lebiere's (A&L's) project of grading connectionism according to the Newell Test as an appropriate means of assessing its utility for cognitive science. In our view, connectionism, unlike ACT-R, is not a specific theory of cognition. It can be used to model nativist, behaviourist, or constructivist theories by modifying parameters with respect to built-in representational and architectural or computational structures. Rather, connectionism is a set of computational and conceptual tools that offer researchers new and precise ways of thinking about and investigating complex emergent behaviour. From this standpoint, if we take the view that theory evaluation in science is best conceived as a comparative affair in which mature theories are evaluated along a number of dimensions to determine which provides the best explanation of the phenomena in question (e.g., Lakatos 1970; Thagard 1992), then connectionism does not offer an appropriate theoretical alternative against which to evaluate ACT-R. Moreover, the current appraisal of connectionism against Newell's criteria actually misses many of the positive applications of connectionist tools in cognitive science research. In developmental psychology, for example, this methodological and conceptual toolbox has been put to use in the service of tackling long-standing issues about the mechanisms responsible for developmental change and, more generally, has supported renewed efforts to construct a genuinely interactional account as a theoretical framework for cognitive development (Elman et al. 1996; Karmiloff-Smith 1992; Newcombe 1998). It has also been successfully used to clarify the fundamental differences between adult neuropsychological patients and children with developmental disorders (Karmiloff-Smith 1997; 1998; Karmiloff-Smith et al. 2002; 2003; Thomas & Karmiloff-Smith 2002) and to model how language acquisition can follow atypical developmental trajectories (Thomas & Karmiloff-Smith 2003).

Connectionist models have been shown to be highly relevant to the concerns of developmental researchers, first, because they offer a valuable means of investigating the necessary *conditions* for development. That is, connectionist models provide concrete demonstrations of how the application of simple, low-level learning algorithms operating on local information can, over developmental time, give rise to high-level emergent cognitive outcomes (Elman et al. 1996; Karmiloff-Smith 1992; Karmiloff-Smith et al. 1998; Plunkett et al. 1997). These demonstrations in turn have forced researchers to revisit assumptions about what can actually be learned as opposed to what has to be prespecified, and to recognize that far more structure is latent in the environmental input and capable of being abstracted by basic learning algorithms than previously imagined.

Concerning assumptions about the nature of the starting state in the developing individual, explorations with connectionist models have been pivotal in clarifying the issue of innateness and identifying a range of potential ways in which innate constraints can be realised (Karmiloff-Smith et al. 1998). As Elman et al. (1996) make clear, despite the current dominance of nativist approaches

https://doi.org/10.1017/S0140525X03310131 Published online by Cambridge University Press

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

https://doi.org/10.1017/S0140525X03310131 Published online by Cambridge University Press