ment theories, with only some weak architectural constraints. Moreover, these languages are computationally universal and thus are equivalent to one another in the sense that one language can simulate the other. How does one evaluate or falsify such universal languages? Are the multiple criteria listed by the authors sufficient to rule out anything at all, or do they simply suggest areas to improve on? The authors' grading scheme is telling in this respect. It only evaluates how an architecture satisfies one criterion better than another criterion, and does not say how to choose between two architectures. One cannot, of course, duck the question merely by choosing an architecture based on the criterion one is interested in explaining. This is precisely the original problem that Newell was trying to address through his multiple criteria.

The authors suggest that timing constraints and memory limitations imply that one cannot only program arbitrary models in ACT-R. But that still leaves room for an infinite variety of models, and ACT-R cannot tell us how to choose between them. To take an analogy to programming languages: It is possible to design an infinite variety of cognitive architectures and implement an infinite variety of models in each one. Can we ever collect enough evidence to be able to choose one over another?

This suggests to me that a cognitive theory must be carefully distinguished from the concrete implementation and the underlying architecture. Just as a programming language can implement any given algorithm, a cognitive architecture can instantiate any cognitive theory (albeit with some variations in time efficiencies). This should not count as evidence for the validity of the architecture itself, any more than good performance of an algorithm should count as evidence for the validity of the programming language. Cognitive science can make better progress by carefully distinguishing the algorithm from the architecture and confining the claims to those parts of the algorithm that are in fact responsible for the results. Consider, for example, ACT-R's theory of past-tense learning by children. More specifically, consider the empirical observation that the exceptions tend to be high-frequency words. A&L attribute this to the fact that only high-frequency words develop enough base-level activation to be retrieved in ACT-R. In more general terms, only high-frequency words provide sufficient training data for the system to be able to learn an exception. How much of this explanation is a result of the particulars of ACT-R theory as opposed to being a necessary consequence of learning in a noisy domain? If any learning system that operates in a noisy environment needs more training data to learn an exception, why should this be counted as evidence for the ACT-R theory? Similar criticisms can be leveled against other cognitive architectures and mechanisms such as SOAR and chunking, connectionism, and backprop.

In other words, even when multiple criteria are used to evaluate a cognitive architecture, there still remains an explanatory gap (or a leap of faith) between the evidence presented and the paradigm used to explain it. To guard against such over-interpretation of the evidence, Ohlsson and Jewett propose "abstract computational models," which are computational models that are designed to test a particular hypothesis without taking a stand on all the details of a cognitive architecture (Ohlsson & Jewett 1997). Similar concerns are expressed by Pat Langley, who argues that the source of explanatory power often lies not in the particular cognitive architecture being advanced but in some other fact such as the choice of features or the problem formulation (Langley 1999). Putting it another way, there are multiple levels of explanations for a phenomenon such as past-tense learning or categorization, including computational theory level, algorithmic level, and implementation level. Computational theory level is concerned with *what* is to be computed, whereas algorithmic level is concerned with *how* (Marr 1982). Cognitive architecture belongs to the implementation level, which is below the algorithmic level. Where the explanatory power of an implementation mostly lies is an open question.

Only by paying careful attention to the different levels of explanations and evaluating them appropriately can we discern the truth. One place to begin is to propose specific hypotheses about the algorithmic structure of the task at hand and evaluate them using a variety of sources of evidence. This may, however, mean that we have to put aside the problem of evaluating cognitive architectures, for now or forever.

ACKNOWLEDGMENTS
I thank Sridhar Mahadevan and Pat Langley for influencing my thinking on this matter and for their comments on an earlier draft.

Cognitive modelling of human temporal reasoning

Alice G. B. ter Meulen
Center for Language and Cognition, University of Groningen, 9700 AS Groningen, The Netherlands. **atm@let.rug.nl** **http://atm.nemit.net**

Abstract: Modelling human reasoning characterizes the fundamental human cognitive capacity to describe our past experience and use it to form expectations as well as plan and direct our future actions. Natural language semantics analyzes dynamic forms of reasoning in which the real-time order determines the temporal relations between the described events, when reported with telic simple past-tense clauses. It provides models of human reasoning that could supplement ACT-R models.

Real-time performance, the second criterion for a human cognitive architecture in Newell (1990), requires the system to operate as fast (or as slow) as humans (target article, sect. 2, Table 1) on any cognitive task. Real time is hence considered a constraint on learning as well as on performance (sect. 5). Although I certainly consider it an advantage of the ACT-R system that it does not rely on artificial assumptions about presentation frequency in the way classical connectionist systems do (Taatgen & Anderson 2002), the limited focus the two systems share on the acquisition of the morphological variability in the simple past-tense inflection in English ignores its obvious common semantic properties, which also must be learned. In this commentary, I propose to include in real-time performance the characteristic human ability to use time effectively when using language to encode information that systematically depends on contextual parameters, such as order of presentation or time of utterance.

Human linguistic competence includes automated processes of temporal reasoning and understanding, evidence of which is presented in our linguistic intuitions regarding the temporal relations that obtain between events described in coherent discourse. The presentation order in which simple past-tense clauses are produced in real time often contains important clues for the correct interpretation. As opposed to the past progressive (*John was leaving*) and the past perfect (*John had left*), the English simple past tense (*John left*) refers to an event that not only precedes the time of utterance but also is temporally located with respect to other events described by prior discourse. The following examples, (1) and (2), show that the order of presentation affects our understanding of what happened.

(1) *John lit a cigarette. He left.*

(2) *John left. He lit a cigarette.*

From (1) we understand that John left after he had lit a cigarette. But (2) makes us understand that the described events occurred in the opposite order. Obviously, the real-time order of presentation in this case determines the temporal relations between the events described. But this is not always so, as we see from examples (3) and (4), where reversing the order of the simple past-tense clauses does not affect the temporal relations between the events.

(3) *John slept for hours. He dreamt of Mary.*

(4) *John dreamt of Mary. He slept for hours.*

Either (3) or (4) makes us understand that John dreamt of Mary while he slept, which is reinforced by the lexical presupposition of dreaming requiring that the dreamer be asleep.

The differences observed between the interpretations of (1)–(4), coincidentally all morphologically strong past-tense inflections, are attributed to the aspectual class of the clauses, which may be telic or atelic (Hinrichs 1986; Partee 1984). Although the compositional characterization of telicity has been a core item on the linguistic research agenda for quite some time, it is generally agreed that in English, clauses that may be modified by durative adverbials, such as *for hours,* are atelic, and clauses that are unacceptable with durative modifiers are telic (ter Meulen 1995; Verkuyl 1996). Temporal precedence effects, which conceptually shift the reference time, are determined by order of presentation of telic clauses in simple past-tense clauses.

Children gradually learn to produce cohesive discourse with simple past-tense clauses, effectively using order of presentation, instead of connecting clauses in their stories with *and then . . . and then. . . .* It depends on their understanding of logical or causal relations between lexical items; for example, dreaming entails sleeping, leaving entails moving elsewhere. It also requires mastering deductive or abductive forms of reasoning, into which neither classical connectionism nor ACT-R have many modelling insights to offer, as Anderson & Lebiere (A&L) readily admit. Reasoning in context and exploiting the dependencies between tense and other indexical features of linguistic expressions cannot be reduced to conditioned correlations between lexical items and concepts, as classical connectionists may want to argue, because it needs a representation of the agent's own information structured information state, as well as a representation of the external domain described by linguistic input and other agents it communicates with. Human understanding of information communicated in ordinary language discourse should, therefore, constitute a core task on the common agenda of cognitive science, testing not only Newell's criteria of real-time performance and natural language, but also adaptive, dynamic, and flexible behavior, as well as knowledge integration and development. Natural language semantics is studying the structured dependencies between context, information, and described domain (Asher et al. 1994; ter Meulen 2000; van Eijck & Kamp 1997). The "Dynamic Turn" in the semantics of both formal-logical, and natural languages has profoundly changed the agenda of the traditional logical systems to require that a dynamic semantics of natural language ideally provides abstract models of our human cognitive capacities of information processing, envisaged in Partee (1997) as the program to "naturalize formal semantics." ACT-R accounts of human cognition may well find it a congenial companion, supplementing its self-proclaimed need for an account of human reasoning.

Real-world behavior as a constraint on the cognitive architecture: Comparing ACT-R and DAC in the Newell Test

Paul F. M. J. Verschure

Institute of Neuroinformatics, University Zürich–Swiss Federal Institute of Technology (ETH), Zürich, 8057, Switzerland. **pfmjv@ini.phys.ethz.ch**
http://www.ini. ethz.ch/~pfmjv

Abstract: The Newell Test is an important step in advancing our understanding of cognition. One critical constraint is missing from this test: A cognitive architecture must be self-contained. ACT-R and connectionism fail on this account. I present an alternative proposal, called Distributed Adaptive Control (DAC), and expose it to the Newell Test with the goal of achieving a clearer specification of the different constraints and their relationships, as proposed by Anderson & Lebiere (A&L).

Anderson & Lebiere (A&L) make the important step to resurrect a number of benchmarks, originally proposed by Newell, which a theory of cognition should satisfy. One benchmark that is missing from this list is that the proposed architecture must be self-contained. *Self-contained* implies that the knowledge of the cognitive system is acquired through an autonomous learning process; that is, its ontology is derived from the interaction between the system and the world. Both ACT-R and classical connectionism do not score well on this constraint. ACT-R fails because it focuses on the use of predefined knowledge in its productions and its recombination by means of chunking. The implementation of its memory structures using artificial neural networks and the inclusion of a subsymbolic/symbolic nomenclature does not address this problem. Classical connectionism fails because it relies on learning rules, for example, backpropagation, that allow the user to compile a predefined input-output mapping into the model (Verschure 1990; 1992). In both cases the models do not tell us how knowledge is acquired in the first place. One could argue that solving this problem of priors is the most fundamental challenge to any candidate theory of cognition (Verschure 1998).

In order to challenge the authors to define more precisely what it takes to satisfy the Newell Test, I present an alternative proposal for a cognitive architecture, called Distributed Adaptive Control (DAC). DAC describes an embodied cognitive architecture implemented by a neuronal system in the context of real-time, real-world behavior. DAC assumes that behavior is organized around three tightly coupled layers of control: reactive, adaptive, and contextual (Fig. 1A). The typical paradigms in which we have developed this architecture are robot equivalents of random foraging tasks (Fig. 1B). It should be emphasized that DAC develops its own domain ontology out of its continuous interaction with the world. Hence, as opposed to ACT-R, DAC is self-contained.

Flexible behavior ("better"). DAC has been shown to organize landmark-based foraging behavior in different types of robots (Verschure et al. 1992; 1996; Verschure & Voegtlin 1998), has been applied to simple games such as tic-tac-toe (Bouvet 2001), has controlled a large scale public exhibit (Eng et al. 2003), and has been shown to be equivalent to an optimal Bayesian interpretation of goal-oriented problem solving (Verschure & Althaus 2003). By satisfying this last constraint, DAC implicitly addresses a wide range of cognitive phenomena (Massaro 1998). This latter constraint argues that our models should attack abstract models describing large repertoires of performance as opposed to single instances of particular behaviors.

Real-time performance ("better"). As opposed to ACT-R, DAC takes real time literally as the time it takes to control real-world behavior. In biologically detailed models, derived from the DAC architecture, of both the sensory (i.e., the learning-dependent changes in receptive field properties of the primary auditory cortex, as reported by Kilgard & Merzenich 1998) and motor aspects (focusing on the cerebellum) of classical conditioning, we have shown that these principles can account for learning performance both in terms of number of trials and in terms of the relevant real-time interstimulus intervals (Sanchez-Montanez et al. 2002; Hofstötter et al. 2002). Hence, these models generalize the hypothesis of DAC towards the neuronal substrate and can account for properties of performance in terms of the underlying neuronal mechanisms. Important here is that temporal properties of behavior are not redescribed in functional terms, which is an underconstrained problem, but directly interpreted in terms of neuronal mechanisms. This illustrates that the benchmarks cannot be interpreted as independent constraints.

Adaptive behavior ("best"). The DAC architecture has been designed in the context of real-world embodied cognition (see also *flexible behavior*). The claim is that only such an approach can account for this constraint. ACT-R is not embodied.

Vast knowledge base (mixed). DAC shows how task-dependent knowledge can be acquired and used to organize behavior and has been applied to a range of tasks (see *flexible behavior*). However, the full neuronal implementation of its structures for short- and long-term memory is not mature enough to make strong statements on its capacity and flexibility (Voegtlin & Verschure 1999). Hence, DAC takes satisfying neuronal constraints as a fundamental benchmark in answering functional challenges. ACT-R seems to stop at a functional interpretation.