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.

https://doi.org/10.1017/S0140525X0342013X Published online by Cambridge University Press

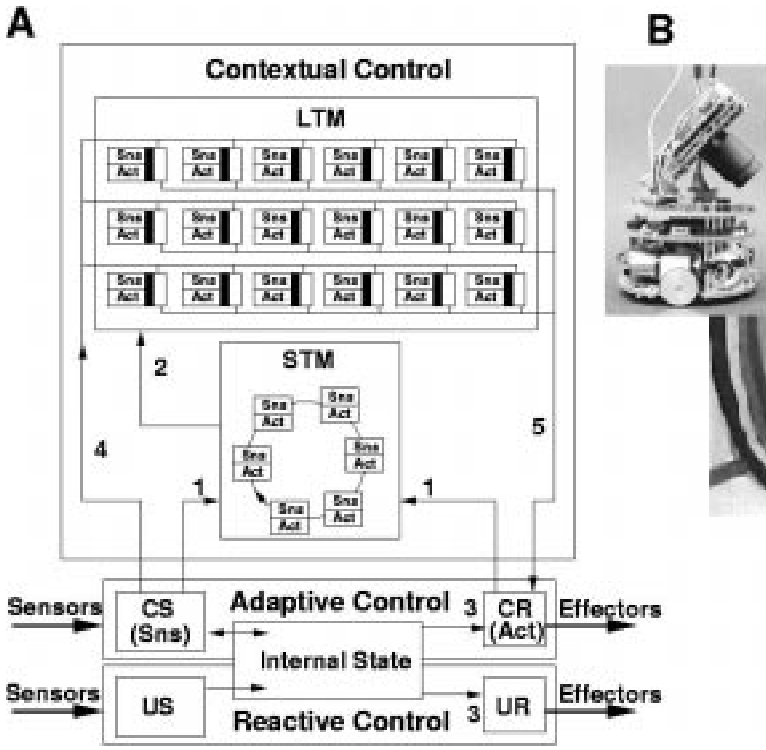

Figure 1 (Verschure). **A.** The DAC architecture. **B.** One example of the application of DAC to robot random foraging using a Khepera micro-robot (K-team, Lausanne). The three layers of DAC each facilitate a progressive structuring of the behavior of the agent. This emergent behavioral organization establishes a non-neuronal feedback loop that guides the perceptual and behavioral learning systems of DAC via behavioral feedback (Verschure et al., 2003).

Dynamic behavior ("best"). DAC has been applied to real-world tasks that include goal conflicts, changing motivational states, and dynamically changing environments, for example, the large-scale exhibition Ada (see *flexible behavior*). In contrast, ACT-R has only been tested on closed problem domains and has not considered the motivational components underlying the organization of dynamic behavior.

Knowledge integration ("better"). DAC has been shown to both acquire the content of its memory structures and to form goal-related recombinations of these representations. Given its Bayesian equivalence, DAC satisfies properties of inference making and induction. However, what is required is a more explicit specification of the experimental data that should be accounted for.

Natural language ("worse"). DAC has not been applied to any form of language acquisition or expression. However, DAC claims that its general learning properties will generalize to language; that is, an explanation of language should emerge from the general principles that underlie the organization of adaptive behavior and not require yet another a priori functional module. In contrast, ACT-R appears to develop in terms of a collection of functionally distinct and independent modules.

Consciousness ("worse"). For now, there is no ambition in the DAC project to attack this phenomenon.

Learning ("best"). DAC was initially conceived to address the behavioral paradigms of classical and operant conditioning. These forms of learning, as opposed to the ones the authors focus on, deal with the problem of autonomous acquisition and expression of knowledge. The biologically detailed models derived from DAC, described above, for instance, account for the phenomenon of blocking central to the Rescorla-Wagner rule of classical conditioning in terms of neuronal mechanisms and not only in functional terms (Hofstötter et al. 2002). This again emphasizes that functional and structural constraints must be satisfied simultaneously and that constraints should be defined around general models, such as the Rescorla-Wagner laws. Moreover, this approach illustrates that a theory of a cognitive architecture will probably be accompanied with a large set of specific derived models that validate a specific subset of its assumptions.

Development ("better"). The DAC architecture interprets development as the progressive involvement of its adaptive and contextual control layers. We have shown that this progression can display stage transitions characteristic for cognitive development (Verschure & Voegtlin 1998). However, the authors should be more precise in specifying what the exact data sets are that should be explained to satisfy this benchmark.

Evolution ("mixed"). Following classic examples of, for example, Pavlov (1928), DAC assumes that cognition arises out of a multilayered architecture that requires a minimum of prior specification. Because the phenomenon of classical conditioning has also been observed in insects (Menzel & Muller 1996), we are currently investigating whether the DAC principles do generalize to insects. Hence, although the results are not in, the claim is that phylogenetic continuity of principles underlying cognition should be evaluated following this comparative approach.

Brain ("better"). As mentioned earlier, the basic principles underlying the adaptive and reactive layers of DAC have been implemented and tested using biophysically and anatomically constrained models. Although the contextual layer makes predictions about the functional properties of neuronal organization, in particular, in relation to the hippocampus, basal ganglia, and prefrontal cortex, these predictions still need to be verified by developing biologically constrained models of these structures. ACT-R seems to stop at finding a correlation between neuronal responses obtained with fMRI measurements and its functional decomposition of cognition. This might not be sufficient. A&L should be congratulated for proposing a common test for theories of cognition and exposing ACT-R to it. The Newell Test in its current form, however, is not mature enough to use it as a gold standard for theories of cognition. This step should be taken in order to advance our understanding of mind, brain, and behavior.

In Figure 1, panel A, the reactive control layer provides a behaving system with a prewired repertoire of reflexes (unconditioned stimuli and responses – US, UR) that enable it to interact with its environment and accomplish simple automatic behaviors. The activation of any reflex, however, also provides cues for learning that are used by the adaptive control layer via representations of internal states. Adaptive control provides the mechanisms for the adaptive classification of sensory events (conditioned stimulus – CS) and the reshaping of responses (conditioned responses – CR) supporting simple tasks, and can be seen as a model of classical conditioning. The sensory and motor representations formed at the level of adaptive control provide the inputs to the contextual control layer that acquires, retains, and expresses sequential representations using systems for short- and long-term memory. The contextual layer describes goal-oriented learning as observed in operant conditioning. Central-processing steps at this level in the architecture are the following: (1) The representations of sensory cues (Sns) and associated motor states (Act) acquired by the adaptive layer are stored in short-term memory (STM) as a segment. (2) If a goal state is reached, that is, a target found or a collision suffered, the contents of STM are retained in long-term memory (LTM) as a sequence. Each segment of LTM consists of a sensori-motor representation (Sns, Act) a trigger unit (black) and a collector unit (white). (3) The reactive and adaptive control layers can still trigger actions and stand in a competitive relation to the contextual control system. (4) Each Sns state of the adaptive layer is matched against those stored in LTM. (5) The collector units of LTM can trigger actions dependent on the biased competition between LTM segments. By modulating dynamic thresholds of each LTM segment, different chaining rules can be implemented.

In panel B of Figure 1, the robot learns to use the color information in the environment, the patches on the floor and the walls, in order to acquire the shortest route between goal locations, that is, light sources (grey circles). The trajectory visualized is generated during a recall task where the light sources are switched off, after learning for about 30 min. The environment measures about 1.5 by 0.8 m; and the robot, about 55 by 30 mm.

A multilevel approach to modeling human cognition

Hongbin Wang,[1] Todd R. Johnson, and Jiajie Zhang

School of Health Information Sciences, University of Texas Health Science Center at Houston, Houston, TX 77030. **hongbin.wang@uth.tmc.edu todd.r.johnson@uth.tmc.edu jiajie.zhang@uth.tmc.edu http://www.shis.uth.tmc.edu**

Abstract: Although we agree with Newell and Anderson & Lebiere (A&L) that a unified theory of cognition is needed to advance cognitive science, we disagree on how to achieve it. A hybrid system can score high in the Newell Test but may not offer a veridical and coherent theory of cognition. A multilevel approach, involving theories at both psychological and brain levels, is suggested.

Newell certainly had a very good reason for being frustrated over the progress toward a scientific understanding of the human mind. The human mind is undoubtedly one of most complex entities in the world. It is systematically shaped by genetic and evolutionary forces; fundamentally constrained by physical and biochemical laws; influenced by cultural, social, and environmental factors; and manifests itself both psychologically and neurophysiologically. Given its inherent complexity and our limited knowledge in each of these aspects, it is conceivable that we may not be able to achieve a thorough understanding of the mind's work for a long time.

While we share Newell's frustration, we doubt that the Newell Test, as proposed in the target article, would offer us relief. On the one hand, the attainability of the test is theoretically questionable. It remains controversial, for example, whether self-awareness and consciousness are computationally implementable (e.g., Penrose 1989; 1996; 1997). This controversy helps to explain why both connectionism and ACT-R were graded "worse" on criterion 8 (self-awareness and consciousness) in the target article. On the other hand, even if we ignore the possible theoretical difficulties, we may still encounter practical problems in developing theories of mind that can pass the test, as we elaborate later.

After evaluating connectionism and ACT-R based on the Newell Test and suggesting that neither was satisfactory on all criteria, the authors Anderson & Lebiere (A&L) go on to recommend some remedies. One major remedy suggested is that we should somehow dissolve the distinctions and join the two approaches close together. Specifically, ACT-R needs to be "more compatible with connectionism," and connectionism needs to be concerned "with more complex tasks and symbolic processing" (sect. 6, para. 3). The authors note that building hybrid systems that can integrate the two approaches is particularly promising (ACT-R itself is already a form of hybrid system). By combining the advantages of different sub-approaches, the authors seem to suggest that hybrid systems would bring us one step closer to a Theory of Mind (ToM) that can score high in the Newell Test.

Unfortunately, there are at least three problems with this hybrid system approach. First, it should be noted that there are two (out of 12) criteria on which both connectionism and ACT-R score worse or worst. They are criterion 8 (self-awareness and consciousness) and criterion 11 (evolution). The simultaneous failure of both approaches on both criteria suggests that simply hybridizing the two approaches might not provide a solution.

Second, what if we develop a theory of self-awareness and an evolutionary ToM, and then hybridize these two theories with the hybrid system we constructed earlier? Does this give us a better ToM? Well, maybe. If doable, it will certainly boost the Newell Test score! But it also induces a paradox. Focusing on isolated and segmented subtheories of mind is what frustrated Newell and motivated the creation of the Newell criteria in the first place. If we first need to develop subtheories to develop high-scoring hybrid systems, we then lose the very point of the Newell Test.

Third, and most important, hybrid systems are artificially assembled systems and thus bear little true psychological and neurophysiological significance. Although we all agree that the human mind is a complex, multilevel construct and involves mechanisms and operations at, among others, both psychological and neuronal networks levels, simply piecing them together is ad hoc and trivializes the problem. A ToM that explains one phenomenon using a neural-network-level mechanism and explains another phenomenon using a rule-based, symbolic-level mechanism may be a convenient hybrid ToM, but is certainly not the *unified* ToM that Newell had wished for (cf. Newell 1990).

In our view, any principled ToM must recognize that the human mind may adopt different mechanisms and follow different laws at different levels. In addition, it is highly unlikely that there exists any simple and linear one-to-one mapping across levels. Penrose, for example, went so far as to hypothesize that there is a noncomputational and nonlocal process called "objective reduction" that connects physics and consciousness (Penrose 1996; see also Woolf & Hameroff 2001). We would argue that a similar nonlinear relationship exists between the neuronal-network-level and the psychological level, and that each level tells a veracious but adequately distinct story of mind. Such a multilevel view is also consistent with both Marr's (1982) and Newell's (1990) conception of multiple-level description of human cognition. Consequently, we should not expect a single architecture, even a hybrid one, to explain all of the phenomena of mind.

We regard both ACT-R and connectionism as celebratory candidates for a ToM, but at different levels. Whereas ACT-R focuses on the symbolic mental structures and processes and offers a psychologically plausible explanation that closely links to empirical behaviors, connectionism adopts subsymbolic neural-based mechanisms and permits a biologically realistic explanation

https://doi.org/10.1017/S0140525X0342013X Published online by Cambridge University Press