judge the worth of complex AI ideas based on armchair contemplation alone. The book itself is easy to understand, though a little verbose. With AI currently in at least a California winter, you have to admire the SOARians' ability to acquire the resources to support 35 people.

Inertia and bull's hide. One of SOAR's weaknesses is the theoretical stinginess of the system. The present SOAR mechanisms are clearly inadequate for unifying cognition and Newell acknowledges this. Some ongoing internal improvements are mentioned, but there is great reluctance to add major new mechanisms or modules. The best explanation is found in Rosenbloom et al. (1990), where they admit that major augmentations to SOAR would probably destabilize the system. Big software systems develop considerable inertia. The SOARians don't want to run the high risk of dismantling SOAR and perhaps being unable to make it work again with major new modules. But if SOAR is going to evolve from a chunking shell into a true artificial mind, it must acquire and *strongly* integrate more features, for example, a language module, inductive learning, and additional reasoning mechanisms and representations.

Another problem is the spotlighting of the "architecture," the shell, rather than the knowledge structures and processes inside the shell that implement a particular cognitive task. Architectures don't do anything. It's the stuff inside that does the work. According to legend, the Phoenician queen Dido founded Carthage by a creative real estate transaction with the local king. For a bargain price the Phoenicians agreed to buy as much land as could be encompassed by a bull's hide. They cut the hide into very thin strips, with which they enclosed a huge area – an early illustration of the power of the block diagram. AI architectures are a device by which busy establishment researchers try to claim vast tracts of AI territory with minimal material and labor.

On putting the cart before the horse: Taking perception seriously in unified theories of cognition

Kim J. Vicente and Alex Kirlik

Center for Human-Machine Systems Research, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332

Electronic mail: *vicente@isye.gatech.edu*

In *Cognition*, Newell provides a scholarly account of the symbolic processing view of mind and makes a convincing argument for the need to develop unified theories of cognition. Although one can hardly question the importance of pursuing this goal, one can certainly debate the best way to achieve it. In our opinion, Newell's approach suffers from a number of problems which stem directly from the fact that the role of perception has been seriously underrepresented in SOAR. Perception should receive more emphasis in unified theories of cognition and questions pertaining to perception must be considered logically prior to the questions pertaining to central cognition that Newell has emphasized in his work.

Newell admits frankly that contemporary theories of cognition have focused almost exclusively on central cognition to the detriment of perception and action (p. 15). In the development of SOAR, he accordingly begins with problem solving, decision making, and routine action. Memory, learning, and skill are considered next. Perception and action appear only third in a priority list of areas to be covered by a unified theory of cognition (p. 15). Newell does not justify this priority scheme, but a letter from Herb Simon to Adriaan de Groot clearly reveals that the choice was not theoretically motivated (cf. Simon 1991, p. 206).

We believe that significant problems arise from giving perception such a low priority. Newell concurs and even admits his guilt:

> The result is that the theory gives up the constraint on central symbolic cognition that these systems [perceptual and motor] could provide. The loss is serious – it assures that theories will . . . never be able to tell the full story about any particular behavior . . . . By describing central cognition first, I am committing this same sin. (pp. 159–60).

In our opinion, however, Newell underestimates the implications of his strategic choice. Not only will the SOAR architecture be incomplete, but the validity of a SOAR model for any task will be limited by the assumptions that have been made concerning perception. Stated bluntly, if you get perception wrong in instantiating the SOAR architecture for a given task, then everything downstream in that model will be wrong as well (cf. Neisser 1987).

How can such a strong claim be substantiated? Note that to build a complete SOAR model for any task, Newell must make strong assumptions about the functionality, if not the processes, of perception and action. Perceptual outputs and motor system inputs must first be defined in order to specify the inputs and outputs, respectively, of the central cognitive systems. While SOAR focuses on modeling the input-output mapping computed by the central systems, such a model can be no better than the description of the inputs and outputs themselves. In other words, the validity of the assumptions about what perception provides and what the motor system accepts places an upper bound on the validity of the theory of how the central systems generate the mapping from perception to action. Improper assumptions about the functionality of perception and action accordingly do not merely render the cognitive model incomplete. Rather, they invalidate the cognitive model before the modeling can even begin.

How reasonable, then, are the assumptions that have gone into the development of SOAR's perceptual system? Given the low priority of perception, it is not surprising to find that SOAR's perceptual and motor systems are immature, as Newell himself points out (p. 160). More specifically, SOAR's perceptual system merely transduces stimulus energies from the world into representations to be acted upon by central cognition. It runs autonomously and cannot be modified by experience. This simple perceptual system is complemented by a set of encoding production rules stored in long-term memory (LTM) whose role is to parse the output from the perceptual system into a representation of the stimulus in task-dependent terms.

This view of perceptual functionality is impoverished and entirely at odds with the theoretical and empirical research that has been generated by Gibson's (1979) ecological approach to perception. [Cf. Ullman: "Against Direct Perception" *BBS* 3(3) 1980.] In this view, people are often able to perceive directly the affordances of the environment – the action possibilities it offers them – through a process of direct attunement, *without* any need for mediating inferential processes. Direct perception is possible because (a) the action-relevant properties of the environment (affordances) are uniquely specified by information (higher-order properties) in the perceptual array, and (b) the observer is able to pick up this information directly by being attuned to it. There is increasing empirical evidence supporting this view of perception (e.g., Bingham et al. 1989; Mark 1987; Savelsbergh et al. 1991; Solomon et al. 1989; Warren & Hannon 1988; Warren & Whang 1987).

What implications does this have for SOAR? First, at the architectural level it indicates that SOAR's perceptual system is inadequate. Second and more important, at the level of models of specific tasks, adopting an incorrect theory of perception has significant repercussions for central cognition. Because of the primitive nature of SOAR's perceptual system, a large number of rules will have to be embedded in SOAR's LTM for any given application in order to derive meaningful task-relevant properties from meaningless impoverished inputs. But if perception is direct as the evidence cited above indicates, these rules (and associated computations) would be entirely superfluous (cf.

Neisser 1987). Thus, SOAR would be modeling functions in central cognition (e.g., encoding productions in LTM) that people actually perform perceptually.

Having said this, it is important to acknowledge that direct perception clearly does not occur in all situations, as Gibson (1977/1982) himself pointed out. In cases when it does not occur, people may indeed be engaging in some types of inferential process. The key point, however, is that Newell and most other cognitive scientists have not even investigated what information people are picking up (i.e., what the inputs to perception are) in specific tasks, and thus, whether or not people are engaging in information processing.

Lest our message be misinterpreted, we reiterate that we wholeheartedly support Newell in his laudable quest for unified theories of cognition. Our major concern is that, in their enthusiasm for taking up Newell's challenge, cognitive scientists will blindly persist in overlooking the very hard work that remains to be done on perception. These issues have been and continue to be downplayed or, even worse, just plain ignored. SOAR is but one culprit, but there are others such as ACT* (Anderson 1983),which also suffers from the limitations outlined above. [See also Anderson: "Is Human Cognition Adaptive?" *BBS* 14(3) 1991.] The message we would like to leave in closing is that a viable unified theory of cognition will not, and cannot, exist until the problems of perception are tackled head on. Moreover, we conjecture that the outcome of such an effort will not be the addition of a mere "subroutine" that can be plugged into current theories of central cognition, but will rather require a substantial reformulation of contemporary cognitive theories.

ACKNOWLEDGMENTS

We would like to thank Steve Casner, Adriaan de Groot, Ken Hammond, Ulric Neisser, and Peter Pirolli for commenting on an earlier version of this review.

Cognition and simulation

N. E. Wetherick

Department of Psychology, University of Aberdeen, King's College, Old Aberdeen AB9 2UB, Scotland

Newell proposes a candidate "unified theory of cognition" – SOAR. The theory has for me a degree of face validity because "problem space" corresponds to an intuited feature of my own problem-solving behaviour and I imagine I am not unique in this respect. I am less happy with what, according to the theory, goes on in the problem spaces. The theory appears to postulate nothing except associations between the values of attributes of objects that constitute "elements" in the working memory (p. 166). Combinations of these elements fire productions drawn from the long-term memory. These productions are sets of previously associated values of attributes of objects, and are fed back as elements into the working memory, where the new set (or one of its subsets) may fire a new production, and so on. Judgments ranging upwards in complexity from simple preference are postulated to occur in the problem spaces but the theory says little about them. They seem to me, however, to involve much more than association based on past experience. Much of Newell's evidence is drawn from current work in cognitive psychology, which is open to the same criticism because it is often prepared to postulate activities of a "central executive" or an "attention director" or a "manipulator" of mental models, without further specification. These may well participate in ongoing cognitive activity and ought to be covered by the theory, but frequently they are not.

My questions arise out of these general concerns:

1. An "element" consists of an "attribute value of an object." In what sense are they values "of an object"? Does Newell subscribe to Gibsonian realism, where the assembly of "values of attributes" into "objects" is given by the structure of the real world outside us? Most perceptual theorists insist, on the contrary, that "objects" are some kind of mental construction from sense data; Newell's "encoding productions" suggest that he may agree with them. If so, how and when are (or were) the necessary associations formed, and how are (or were) they validated? (Since the sense organs, on this view, furnish only "values of attributes.")

2. Elements may be negative. I take this to mean that while elements *A* and *B* may fire a production, they do so only if *C* is not also present. How does the system establish that *C* is not present? Not by exhaustive search, since the working memory does not have a fixed capacity. Can productions be inhibited by elements? I do not recall this possibility having been discussed.

3. I consider here the special case of syllogistic reasoning since it is an area with which I am familiar, but the line of argument could, I think, be pursued in other areas too.

Newell begins by postulating an "additional hypothesis beyond the SOAR architecture" (p. 396). Humans use "annotated models" in syllogism solving, as described by Johnson-Laird (1983). There is no conflict between model theory and problem space theory. He goes on to report an elaborate simulation in SOAR of the task set by Johnson-Laird and Bara (1984) to 20 Italian subjects, which generated a close match to their results (Fig. 7-11, p. 404). He does not enquire into the validity of the results.

Inspection of Fig. 7-11 shows that

(i) No subject ever proposed an *A* conclusion unless both premises were *A*'s.

(ii) No subject ever proposed a correct subaltern conclusion (i.e., *I* for *A*, or *O* for *E*).

(iii) Only about 6% of the conclusions proposed were "nonmatching" (i.e., different in logical form from both premises). Of the 27 syllogisms having valid conclusions, 17 require a nonmatching conclusion – over 60%!

It appears that the subjects were unwilling to commit themselves to an *A* conclusion except in very restricted circumstances. Normally they proposed a conclusion of the same logical form as the non-*A* premise, or as either one of the premises if neither was an *A*. For example, although 18 out of 20 subjects proposed an *I* conclusion to *IA* (syllogistic Fig. 1) which is correct, 18 also proposed an *I* conclusion to *II* (syllogistic Fig. 1) and 13 to *AI* (syllogistic Fig. 1), both of which are incorrect. It appears that well over half the subjects must have proposed *I* in all three cases. This behaviour has nothing to do with logic. The importance of this and other similar facts is that all Johnson-Laird's one-model syllogisms ($n = 10$) would have been solved correctly, since the correct conclusion is of the same logical form as the non-*A* premise, whereas all the three-model syllogisms would have been failed, because the correct conclusion is not of the same logical form as either premise. This is more or less what happened. (Few of the totals add up to 20, so there must also have been other wrong conclusions which are not listed.)

Whatever it was that these subjects were doing (it does not appear to have been syllogistic reasoning), Newell has simulated it successfully. Subjects can show that they are trying to do syllogistic reasoning by proposing correct conclusions to three-model syllogisms but none of the Italian sample did so (see Johnson-Laird, 1983, Table 5.6), although they made few errors on one-model syllogisms. In Johnson-Laird's earlier (American) sample (see Table 5.5), the first two or three subjects may have been reasoning syllogistically. I have shown elsewhere (Wetherick 1989; Wetherick & Gilhooly 1990) that many subjects in this type of experiment arrive at their conclusions by the matching procedure described above – many, but not all. In my studies there are always one or two subjects who reason syllogistically without having been taught to do so. (All of them can of course be taught; not having been taught they have no way of knowing that the matching procedure is wrong – after all, it gives the right answer to one-model problems!)