Computational Intelligence, Volume 9, Number 4, 1993

IS A PICTURE REALLY WORTH A THOUSAND WORDS?

MICHAEL J. TARR
Department of Psychology, Yale University, New Haven, CT 06520-7447, U.S.A.

1. WHAT IS IMAGERY?

If the author wishes to bring the imagery debate to the artificial intelligence (AI) community, I believe it is important that we begin with a better definition of what one means by "imagery." Of course, almost all humans have an intuition about what it is like to imagine something; but it is doubtful that such experiences inform us very much about the format or processes underlying imagery. However, this does not lead me to dismiss empirical methods as the primary means for understanding imagery. Cognitive psychologists are constantly facing this challenge—our goal is to develop paradigms that circumvent or otherwise control for subjects' intuitions, but result in measurable behaviors that reveal properties of the underlying mechanisms and representations of cognition. Thus, at least for a first pass, I wish to retain the notion of imagery as a uniquely human phenomenon and suggest that it should be studied as such.

In some ways this may seem to be a highly unsatisfactory assumption, for we are left with a definition of imagery that is a tautology: imagery is simply those instances in which humans imagine. However, there is more to this definition than meets the eye. First, because this definition necessitates a *conscious* being to experience the image, it immediately narrows the meaning to encompass only the human (and possibly other animals') experiences of imagery, while eliminating instances in which a so-called image is processed by a machine (at least until we develop conscious machines). Second, because this definition is modality neutral, it accommodates *any* experience in which we encounter images, including imagery in all of our senses, as well as nonsensory thoughts and desires. This issue is raised, but is rightly tempered by the fact that most studies of imagery, as well as the debate itself, have centered on visual imagery. However, while the author takes this as a basis for quickly shifting from imagery to spatial representations,[1] this emphasis obscures the essential point of the imagery debate—whether or not human information processing beyond early perceptual input utilizes modality-specific representations. Thus, the imagery debate was not about whether humans have visual representations (of course they do), but whether such representations are retained and play some functional role in our cognition. This point is crucial to current theories of imagery—rather than simply stating that we have representations that "are like" structures in the world, it has been argued that we retain perceptual input in a form that allows us to reinstantiate the prior perceptual processing of this information by activating some of the same structures used for the original perception. Thus, it is not the fact that images are derived from percepts that makes them images, but rather that they can function as percepts at some later time.

[1]The paper adopts Farah *et al.*'s (1988) distinction between *visual* and *spatial* representations, but these terms are misnomers. The essential property of visual imagery is that all images are spatial representations, having at least a two-dimensional coding of information isomorphic with its arrangement in the physical world. Thus, I am using the term "spatial" in a broader sense, and reserve the terms *intensity based* and *structural* to denote the distinction between two types of spatial representations.

Indeed, it is this conception of imagery that has been adopted by modern cognitive and neuropsychology (Farah 1985).

In contrast, in invoking the concept of imagery, I believe that the author means something significantly different from the above portrayal. First, by necessity, within the computational model there is no reference to consciousness; rather, images are acquired by sensors or input by an operator and then processed according to a prespecified set of instructions. Second, the model is restricted to spatial representations (either intensity based or structural)—it is not designed to accommodate representations that code information in other formats, for instance, air pressure over time (sound). Third, the model has no notion comparable to the reinstantiation of processing within perceptual mechanisms—instead, there are a suite of "inspection" processes that may be invoked to recode or transform the input. Thus, the model may be characterized more accurately as a method for encoding spatial information in conjunction with mechanisms for explicitly deriving qualitative properties from such information. While this characterization has something in common with the representations and processes proposed in theories of human visual imagery, it is neither a general account of imagery, nor a specific emulation of particular aspects of our imagery system. Therefore, its utility must lie elsewhere.

2. GOALS OF COMPUTATIONAL THEORY

There are three possible goals for proposing a computational theory. First, a computational theory might be intended as a simulation and possibly as an explanation of human cognitive capacities. For instance, Marr (1982) set forth a wide range of computational models explicitly designed to explain particular aspects of human vision. Second, a computational theory might serve as part of an AI system, taking input and producing output, either in the form of actions or computed information. For instance, AI systems have been developed to configure computer systems, play chess, and diagnose symptoms of diseases. In such instances there is little attempt to mimic or otherwise explain human behavior (although information processing strategies derived from human performance are often adopted)—the goal is successful task completion. Third, a computational theory might serve as a tool for investigating the implications of a specific format for representation and specific processing mechanisms over such representations. This final goal is tacitly independent of either human or AI performance. However, in practice, this is rarely the case—such test-beds are designed either as simulations of biological systems or as precursors to functional AI systems.

In considering the model in the context of these three goals, it becomes apparent that something of a hybrid is intended. First, while the author is careful to distance the model from any attempt to account for extant cognitive and neuropsychological data on visual imagery, because she wishes to bring the imagery debate to AI, much of the model has the flavor of psychological theory (Kosslyn 1980, in particular). Second, it is clear that the model is not intended as an AI system *per se*—although one could envision an AI system that made use of these representations and mechanisms as part of a method for drawing inferences from spatial input. Third, and perhaps most importantly, the model is intended "to provide insight into what representation schemes and reasoning strategies may be most appropriate for problems that involve imagery"—an aim consistent with a goal of investigating the computational properties of representations and processes. Indeed, in light of the implicit claim that the AI community has not paid much attention to or otherwise utilized spatial representations, this is an important contribution.

The flaw with this argument may be that the AI community *has not* ignored spatial representations. Indeed, within many branches of AI there has been extensive use of spatial representations for reasoning about navigation, scene understanding, and many physical properties of the external world (Chen 1990). Beyond these examples, computer vision, certainly a part of the AI community, has concentrated almost entirely on the representation and processing of spatial information. Moreover, as a psychologist who studies visual cognition, I consider the accomplishments of these areas of AI to be significant—contributing both to a better understanding of the computational properties of spatial representations and to better theories of visual cognition in humans (Pinker 1985). Thus, when a representational format for spatial information is proposed and is cited as providing computational advantages over other informationally equivalent representations, the author is stating something that certain AI researchers have been pursuing for many years. Of course, it is doubtful that the author is unaware of these endeavors; therefore a more liberal interpretation of her thesis is that she is advocating the introduction of spatial representations to other areas of AI—presumably because such areas, for instance, expert systems, have not previously considered them. However, it is my thesis that the absence of spatial representations from these areas has more to do with the sophistication with which past AI researchers have investigated spatial representations, than with their lack of acquaintance with this topic.

3. WHAT IS IMAGERY GOOD FOR?

Reasons for the absence of imagery from some areas of AI may lie in what appear to be the most effective uses of spatial representations. For instance, most of the examples presented in the target article for demonstrating the computational efficiency of spatial representations are tied to reasoning about spatial properties of the world. Of course, it is almost a given that spatial representations isomorphic to the scene will offer a more efficient means for inferring qualitative information about size, shape, structure, etc. than will recodings into alternative representations that do not preserve spatial properties. Indeed, this is one of the primary reasons that many psychologists have been proponents of the existence of imagery—reasoning about the visual world is best done with visual representations (witness the incredibly convoluted and Rube Goldbergesque nature of the nonspatial theories of imagery proposed by propositionalists). Other examples of the supposed advantage of spatial representations are misleading in another fashion. For instance, the example in which the relationships between the heights of individuals are readily inferred when presented spatially, but would be somewhat more difficult to infer if presented propositionally, relies not on any intrinsic efficiency for spatial representations, but on the fact that human cognition apparently has mechanisms for comparing the magnitudes of shapes in parallel (Ullman 1984), but no comparable mechanisms for propositions. Therefore, it is possible that, as a field, the AI community *has* considered spatial representations—it is just that some branches have found them more useful than others.

Thus, while the author's objective is a reasonable one, I suggest that it is unwarranted on several counts. First, it does not provide a theory of human imagery; hence it is of little use for modeling biological information processing. Second, some within the AI community have considered spatial representations, but have discarded them in favor of representations more appropriate to the problem domains at hand. Third, it is preaching to the converted, in that many within the AI community have adopted spatial representations for precisely the reasons of computational efficiency advocated here.

ACKNOWLEDGMENT

The support of the Air Force Office of Scientific Research, contract number F49620-91-J-0169, is gratefully acknowledged.

REFERENCES

CHEN, S. *Editor.* 1990. Advances in spatial reasoning. Ablex. Norwood, NJ.

FARAH, M. J. 1985. Psychophysical evidence for a shared representational medium for mental images and percepts. Journal of Experimental Psychology: General. **114**:91–103.

FARAH, M. J., K. M. HAMMOND, D. LEVINE, and R. CALVANIO. 1988. Visual and spatial mental imagery: dissociable systems of representation. Cognitive Psychology, **20**:439–462.

KOSSLYN, S. M. 1980. Image and mind. Harvard University Press, Cambridge, MA.

MARR, D. 1982. Vision: a computational investigation into the human representation and processing of visual information. Freeman, San Francisco.

PINKER, S., *Editor.* 1985. Visual cognition. MIT Press, Cambridge, MA.

ULLMAN, S. 1984. Visual routines. Cognition, **18**:97–159.