

A Review of *Sketches of Thought*

B. Chandrasekaran

Most practicing AI researchers take certain things for granted as part of the background assumptions of their craft. That intelligence is a form of information processing and that the framework of modern digital computers provides pretty much all that is needed for representing and processing information for doing AI are two of the most foundational of such assumptions. Turing (1950) explicitly articulated this idea in the late 1940s, and later Newell and Simon (1976) proposed the physical symbol system hypothesis (PSSH) as a newer form of the same set of intuitions about the relation between computation and thinking. In this tradition, the computational approach is not just one way of making intelligent systems, but representing and processing information within the computational framework is necessary for intelligence as a process, wherever it is implemented. Philosophy of mind, linguistics, and cognitive science all take the computation-over-representation hypothesis as central to understanding human cognitive phenomena as well. The language of thought (LOT) hypothesis, of which Fodor (1975) has given the most well-known exposition, is a variant of the computational hypothesis in AI. LOT holds that underlying thinking is a medium that has the properties of formal symbolic languages that we are familiar with in computer science. There is such a close connection between our current notions of thinking, cognition, intelligence, and so on, on the one hand, and computation and representation, on the other, that for most AI practitioners, it is hard to imagine that this basic hypothesis could be questioned.

However, this hypothesis has increasingly been under attack from various sources in the last decade. First came connectionism with its challenge to what has come to be known as *symbolic representation*, that is, representation using one form or other of digital computer languages. Connectionism was followed in quick succession by proposals for other nonsymbolic representations, such as dynamical systems (see Port and van Gelder [1995] for a recent collection of articles). Within AI, Brooks (1986) claimed to be producing intelligent behavior without any representation at all. There has also been a debate within cognitive science and AI about exactly what the implications of the situated cognition ideas were for the

Sketches of Thought, Vinod Goel,
The MIT Press, Cambridge, Massachusetts, 1995, 279 pp., ISBN 0-262-07163-0.

representation hypothesis. Several researchers from this movement claimed that representations are not really processed to produce intelligent behavior as much as representations are constructed as part of intelligent behavior, and that much of AI, including the PSSH, had misconstrued the nature of cognition (Cognitive Science 1993). Simon himself (Cognitive Science 1993) claimed that, Brooks's own claims notwithstanding, his robots were actually having and using representations and that his work fully satisfied the requirements of the PSSH.

There was also trouble for the symbolic hypothesis from another

source. This was the idea that a good part of intelligent behavior was actually based on iconic or pictorial representations inside the head, which, in some tellings, were incompatible with the traditional symbolic approach but in others were quite compatible. For AI, this debate opened up the issue of exactly what is meant by pictorial representations and when they were useful (see Glasgow, Narayanan, and Chandrasekaran [1995] for a recent collection of articles).

Most AI researchers ignore all this turbulence regarding the foundational hypothesis of computation over representations—they go on formulating technical problems that are amenable to progress using the traditional hypotheses. However, many in the field see these issues as important, not only for abstract philosophical reasons but for quite pragmatic reasons—perhaps there are better ways to build effective artifacts.

What becomes clear to anyone who wants to make sense of the issues is how confusing and confused the foundational ideas are. Ideas that seem reasonably straightforward at first blush—representation, for example—turn out to be quite elusive once one tries to pin them down. To illustrate, the difference between so-called propositional and pictorial representations, an idea that appears to be simple intuition at first, gets mired in complexities when one begins to formalize these representations.

This is where Vinod Goel's book, *Sketches of Thought*, comes in. Goel's main goal is to propose and defend the hypothesis that the ideas about representation that underlie much current research in AI and cognitive science do not do justice to the full range of representational powers that the human cognitive system displays. I think that the foundational issues that the book deals with are important for AI researchers. I try in this review not to engage the book at the level of its detailed arguments because that would call for a much longer review. I attempt to give an outline of its arguments and positions in its own terms, so that the reader gets a general idea of why I think the book is an important and useful contribution.

Goel devotes the first part of the book to an excellent discussion of the idea of representation in models of thought. He starts with cognitive science's commitment to explaining thinking as symbol manipulation, specifically, symbol processing by computation. He describes the subtle distinctions between different versions of computational theories in AI and cognitive science, in particular between the PSSH in AI and the LOT hypothesis in cognitive science.

He identifies a set of properties—the so-called CTM properties—that the computational theory of mind, as used in cognitive science and AI, is necessarily committed to. Among them are syntactic and semantic differentiation, causally efficacious syntax, right causal connections of the underlying physical states, and unambiguity. These properties need a little explanation. Consider the simple example of marks on paper representing symbols in some symbol system. Different marks on paper might all be versions of the letter A. The marks are tokens, and the type they denote is the symbol A. A computer reading a mark is in some physical state. There is an equivalence class of physical states (corresponding to all the possible marks for this symbol) that correspond to this symbol. This equivalence class of physical states, when the device is a computer and viewed as one, is also a computational state.

The syntactic constraints in the list of CTM properties refer to the marks side. *Causally efficacious syntax* means that the operation of the computational system is causally dependent on the marks—different marks, potentially different state transitions. The *syntactic disjointness criterion* stipulates that the equivalence class of physical states—computational states for each type of marks must be disjoint. *Syntactic differentiation* essentially means we have distinct, not continuous, physical-state equivalence classes. *Unambiguity* means that each equivalence class of physical states—computational states always denotes the same symbol whenever it occurs during the process. Semantic differentiation appears to be really as much a constraint on the world to

which the computations actually refer as it is a constraint on the computational system. Suppose computational state c_s refers to apple, and c_s' refers to orange. Given an apple, we should be able to say about it that it is not an orange, and vice versa. This is a constraint on the world because it says something about how differentiable the semantic categories in the world are (with respect to some perceptual repertoire). It is also a constraint on the computational framework because we want to make sure that the organization of computational states makes use of the differentiability available in the world.

As mentioned, computational models of the sort AI pursues satisfy these requirements. Goel argues that systems with CTM properties are appropriate for well-structured problems but fall short for some ill-structured problem spaces. For these kind of problems, the disjointness and unambiguity properties of CTM systems actually stand in the way. He makes his case for this claim by describing a set of experiments that he ran on designers and the representations they used. First, a good deal of their representations are pictorial in nature, but he doesn't make the commonly made claim that CTM systems are good for propositional (or linguistic) representations and poor for pictorial ones. Instead, he makes a subtler distinction between sketches, which have quite a bit of vagueness, ambiguity, and nondisjointness about them, and representations, pictorial or otherwise, which, in fact, satisfy the CTM requirements. He concludes from his experiments that designers use sketches in an earlier, more conceptual stage of the representations, but their representations tend to get closer and closer to CTM properties as the design document gets closer to delivery to the customer. The diagrams and descriptions in the delivered design document have to be unambiguous and disjoint syntactically and semantically disjoint as well.

Before he gets to the description of his experiments with designers, he takes an interesting detour in which he presents his version of Nelson

Goodman's (1976) analysis of symbol systems. This detour is motivated by a need to understand the distinction between so-called pictorial and propositional representations. He finds most of the previous writing on this subject unsatisfactory. For Goel, Goodman's symbol system framework offers an analytic framework for rethinking the entire issue. The most important aspect of Goodman's system is its identification of three modes of reference, reference being the issue of how a symbol refers to objects or phenomena in the world. The usual mode of reference is denotation; a symbol denotes an object in the world. Nelson adds two: (1) exemplification and (2) expression. A tailor's swatch exemplifies certain properties of the fabric—for example, color and texture—by actually having these properties. However, the manufacturer's name printed on the swatch refers to the manufacturer by denoting. A swatch might represent wealth: Someone who wears a suit made of this fabric might be taken to be rich by someone who looks at the wearer. However, the "expensive-lookingness" of the swatch is a metaphorical form of exemplification. Goodman names this kind of reference "referring by expressing." Goodman felt, and Goel agrees, that we would never be able to make cognitive science give useful accounts of cognitive activity such as painting and music without the richer vocabulary of reference available from the Goodman system. Goel also suggests that exemplification is an especially important form of reference for understanding diagrammatic representations. CTM systems seem to be restricted to denotation as the main basis of reference, with some hope that exemplification might eventually be included.

Goodman uses five criteria to categorize symbol systems: Four of the criteria are somewhat modified versions of the syntactic and semantic disjointness and finite-differentiation criteria that we considered earlier, with a new one, semantic unambiguity, constituting the fifth criterion. Goel singles out 3 of the 32 categories that result from these criteria: (1) no

tational systems, examples of which are artificial languages such as a zip code and a musical score; (2) discursive languages, with natural languages and predicate calculus as examples; and, (3) nonnotational systems, such as painter's sketches, paintings, and sculptures. Notational systems meet all five criteria. A discursive language only meets the syntactic disjointness and finite-differentiation criteria but fail all the semantic ones. Nonnotational systems meet none of the criteria. Although the relationship of CTM criteria to the Goodman criteria are clear, from Goel's discussion, it is not clear where CTM systems fit into the classification. He seems, however, to want to focus mainly on nonnotational systems because he wants to show that some of the external-representation systems used by designers are of this type.

Goel makes what appears to me unsatisfactory distinctions between design and nondesign problems. Any categorization in which a host of problem types are herded under the category non-X should be suspicious as a way of carving up the world in an illuminating way. Even though he spends a lot of time on this issue, I don't think it is central to his basic argument. All he needs to show is the existence of one class of problems where humans use representations that don't have CTM properties. In any case, his experiments demonstrate to him that designers' sketches, especially in the conceptual design stage, are nonnotational, that is, non-CTM. This aspect is not a bug but a feature: The non-CTM properties play a helpful role in the early stages of design. For example, Goel notes, "The failure of the symbol system to be syntactically disjoint gives the marks a degree of coarseness by allowing marks to belong to many different characters. This is a necessary condition for remaining noncommittal about the character....[Not being syntactically finitely differentiable] gives the marks a degree of fine-grainedness by making every distinction count as a different character. This reduction in distance between characters is necessary to help trans-

form one character into another (p. 191)," and so on. His characterization of sketches suggests that sketches need not be restricted to the pictorial domain. Any representation that has inherent vagueness, ambiguity, and nondisjointness but that is nevertheless useful in some modeling and thinking tasks is a sketch.

Having argued that non-CTM symbol systems are needed as external symbol systems at least for some problems, he claims, as a final move, that the internal symbol systems can't all be CTM-like either. Simply, there is no mapping that would go from a CTM-like internal representation to non-CTM external representations, and vice versa. This argument is somewhat complicated, but in essence, it is as follows: Interpreting syntactically undifferentiated or semantically ambiguous marks as equivalence classes requires information that is not present in the marks alone but in the symbol system as a whole; that is, just the mapping alone calls for some sort of a cognitive agent, begging the question of explaining cognition. Lest connectionists think that somehow their proposals for internal representation can then take center stage, Goel argues that connectionism is even less capable in this respect than CTM representations.

Where does all this leave the cognitive scientist and the AI researcher? Goel himself only addresses the cognitive scientist in this book. He says that his arguments speak not against computationalism as such but against the adequacy of a specific type of it, namely, the CTM models. Perhaps there is an account of computation that will enable us to give a computational account of cognition independent of the CTM properties. He suggests that Smith (1996) in his recent book is attempting something of this sort.

What about AI? Well, there are two kinds of AI: There is the AI that thinks that current ideas of computation and computational models are basically all that is needed to enable us to build machines that have cognitive and robotic behaviors essentially coextensive with those of humans.

This wing of AI will not take kindly to the conclusions in this book, although even they can learn quite a bit about the foundational issues in representation. The AI that wants to keep on building smarter and smarter artifacts using current ideas while seeking ways to extend them—that is, the wing of AI that regards itself as having a commitment to the problem rather than to a particular method of solution—will be less likely to object to the book's conclusions.

References

- Brooks, R. A. 1986. Intelligence without Representation. *Artificial Intelligence* 47(1–3): 139–159.
- Cognitive Science. 1993. *Cognitive Science* (Special Issue on Situated Action) 17(1).
- Fodor, J. A. 1975. *The Language of Thought*. Cambridge, Mass.: Harvard University Press.
- Glasgow, J.; Narayanan, N. H.; and Chandrasekaran, B., eds. 1995. *Diagrammatic Reasoning: Cognitive and Computational Perspectives*. Cambridge, Mass.: The MIT Press.
- Goodman, N. 1976. *Languages of the Art: An Approach to a Theory of Symbols*. Indianapolis, Ind.: Hackett.
- Newell, A., and Simon, H. A. 1976. Computer Science as Empirical Enquiry: Symbols and Search. *Communications of the ACM* 19(3): 113–126.
- Port, R. F., and van Gelder, T., eds. 1995. *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, Mass.: MIT Press.
- Smith, B. C. 1996. *On the Origin of Objects*. Cambridge, Mass.: MIT Press.
- Turing, A. M. 1950. Computing Machinery and Intelligence. *Mind* N.S.59: 433–460.



B. Chandrasekaran is director of the Laboratory for AI Research at The Ohio State University. He is a fellow of the American Association for Artificial Intelligence, the Association of Computing Machinery, and the Institute of Electrical and Electronics Engineers. His current interests are in functional and causal reasoning, design problem solving, knowledge systems, diagrammatic reasoning, and foundations of AI and cognitive science. His e-mail address is chandra@cis.ohio-state.edu.