

B. Chandrasekaran, "Multimodal Representations as Basis for Cognitive Architecture: Making Perception More Central to Intelligent Behavior," in *Intelligent Information Processing*, Mark Musen, Bernd Neumann, & Rudi Studer, Eds., IFIP International Federation for Information Processing Series, Vol. 93. Dordrecht: Kluwer Academic Publishers, pp: 13-16, 2002. ISBN 1-4020-7171-X.

Multimodal Representations as Basis for Cognitive Architecture

Making Perception More Central to Intelligent Behavior

B. Chandrasekaran

Dept. of Computer & Inf. Science

The Ohio State University

Columbus, OH, 43210, USA

Abstract: In this talk, I outline a view of "cognitive state" as fundamentally multi-modal, i.e., as an integrated and interlinked collection of "images" in various modalities: the perceptual ones, and the kinesthetic and conceptual modalities. Thinking, problem solving, reasoning, etc. are best viewed as sequences of such states, in which there is no intrinsically preferred mode. Representational elements in one mode invoke elements in other modes. The external world also at various points contributes elements to one mode or another. Perception and imagination are more continuous in this view than in the traditional views. In recent years, there has been much interest in the notion of "mental images." However, the focus in this stream of research has been on a very special class of mental images, namely visual ones. The proposed view is an extension and generalization of this notion, not only to other perceptual modalities, but also to kinesthetic and conceptual modalities. I think the proposed view of the essential nature of the mental state opens up new ways of thinking about cognitive architecture, and also suggests new ways of building smart machines. I'll outline why I think so.

Key words: Cognitive architecture, multimodal representation, agent architecture.

1. EXTENDED ABSTRACT

In this talk, I outline a view of "cognitive state" as fundamentally multi-modal, i.e., as an integrated and interlinked collection of "images" in various

modalities: the perceptual ones, and the kinesthetic and conceptual modalities. Thinking, problem solving, reasoning, etc. are best viewed as sequences of such states, in which there is no intrinsically preferred mode. Representational elements in one mode invoke elements in other modes. The external world also at various points contributes elements to one mode or another. Perception and imagination are more continuous in this view than in the traditional views. In recent years, there has been much interest in the notion of "mental images." However, the focus in this stream of research has been on a very special class of mental images, namely visual ones. The proposed view is an extension and generalization of this notion, not only to other perceptual modalities, but also to kinesthetic and conceptual modalities. I think the proposed view of the essential nature of the mental state opens up new ways of thinking about cognitive architecture, and also suggests new ways of building smart machines. I'll outline why I think so.

My argument will proceed in a number of steps:

(i) The standard view in both AI and Cognitive Science is that "thought" is essentially like a sentence in natural language, i.e., it is either a proposition or an "attitude" (like want P) towards a proposition. The Language of Thought Hypothesis makes this claim in a formal way. All the AI knowledge representation approaches, whether or not their authors think of them as being in the so-called "logic framework," are essentially symbol structures in which the symbols stand for objects and relations in some world.

(ii) I propose that this view of the cognitive state is too narrow, and that having a thought is typically not merely entertaining a proposition or some attitude to it, but having perceptual experiences as well. This claim is not simply of philosophical interest, but I claim that it opens up ways of thinking about building robots that make use of their experience in interesting new ways.

(iii) Let us start with the simple, but common, example of reasoning with external diagrams. An essential role of diagrams in many cases is that the problem solver "sees" some relationship in the diagram; without the diagram, he would, at best, need to construct a series of inferences to arrive at the same piece of information. The relevant point for me here is that the representation of the problem state is distributed between the external diagram and internal representations inside the head of the problem solver. It is also bi-modal - some of the representation is perceptual and some propositional.

(iv) Now suppose there is no external diagram, but the problem solver has to imagine the situation diagrammatically. Think of the example, "Imagine taking one step forward, one step to the right and one step back.

Where are you with respect to the starting point?". Controversy about mental images notwithstanding, the mental image is playing the same functional role in the solution of the problem in this case as the external diagram. The problem state is still bi-modal, but entirely inside the head of the problem solver.

As an aside, a significant aspect of the debate over mental imagery is over the conundrum of how mental images can really be images, since images, in order to be seen, have to be processed by a perception system, and there are no eyes inside the head. I'll briefly outline an earlier proposed of mine as a solution to this paradox.

(v) The next step in the argument is to generalize the representation to multiple modalities. As far as external representations go, in addition to images, we can have three-dimensional models which we can take in visually or by touch. Sound can mark locations and carry specific information. Similarly, there is no reason to restrict the notion of an internal image to the visual modality. In principle, internal images are possible in all modalities of perceptual and kinesthetic experience. In addition to imagining visual situations, we "hear" tunes; we can imagine moving our bodies in various ways, and so on. Clearly some modalities are stronger than others, and people differ in the kinds of modalities their internal images are rich and complex. (Beethoven famously continued to compose music even after he became deaf.) So at this point, we can begin to think of a multi-modal state, spanning all modalities, perceptual, kinesthetic and of course conceptual. The state can be distributed between the inner and outer worlds of the agent.

(vi) I then propose that the cognitive architecture in which the cognitive state is multimodal in the sense above, and describe a number of multi-modal mechanisms that can cause changes of states. When combined with a goal-directed engine such as Soar, we have the basics of a framework for problem solving in which perception is not simply restricted to giving information about the external world, but can support thinking more directly by supporting exploration of multi-modal problem spaces.

(vii) So far it might appear as if the goal is to give a computational account of cognitive states in humans or even animals, but without any particular reason to believe that this is a good architecture for biological agents, or for robots. I will argue for the computational advantages of such an architecture, and propose that it is actually a good idea for natural or artificial agents. I'll conclude with reasons why this way of thinking is a new way of thinking about integrated robots.

(viii) I'll end with an account of technical issues that need to be solved to make progress in realizing this vision. A key challenge is technical solutions to the problem of representing general images in such a way that they are

structured so that they can be composed in various ways, and at the same time can support information extraction, much as we get information from external and internal images.

Acknowledgement: Prepared through participation in the Advanced Decision Architectures Collaborative Technology Alliance sponsored by the U.S. Army Research Laboratory under Cooperative Agreement DAAD19-01-2-0009.