Schemas vs. Symbols: A View from the 90s

Michael A. ARBIB*

University of California at San Diego, Psychology

University of Southern California, Professor Emeritus

Journal of Knowledge Structures & Systems

August 2021
Vol.: 2 Issue: 1 Pages: 68-74

Abstract

Thirty years ago, I elaborated on a position that could be seen as a compromise between an "extreme," symbol-based AI, and a "neurochemical reductionism" in AI. The present article recalls aspects of the espoused framework of schema theory that, it suggested, could provide a better bridge from human psychology to brain theory than that offered by the symbol systems of Newell and Simon.

Key words: Schemas; Schema Theory; Schema-Based AI; Brain Theory; Knowledge; Symbol Systems

1 Introduction

I had the pleasure of meeting Allen Newell and Herbert Simon several times from 1963 to the early 1990s, and much of our conversation was addressed to comparing their symbolic approach to AI with my approach (blending neural networks and schema theory) to modeling brains. These conversations were the basis for my extended review (Arbib, 1993) of Allen Newell's book *Unified Theories of Cognition* (Newell, 1990). In lieu of a Commentary based directly on the text of Luis Augusto's "From symbols to knowledge systems: A. Newell and H. A. Simon's contribution to symbolic AI" (Augusto, 2021), I instead offer lightly edited portions of the 1993 review (too long to reproduce here) that invite the reader to consider the brain-theoretic alternative to symbolic AI.

* ✉ arbib@usc.edu

2 Three Positions in AI

Newell's approach to cognitive science was close to the following "extreme AI" position:

> POSITION (A). Cognitive tasks rest on a set of basic processes, such as pattern recognition, search, memory, and inference, whose properties and interactions can be characterized in an abstract way independent of implementation. In particular, properties and interactions that we understand by implementing them on serial symbol processors are equally valid for the human mind.

Many AI workers (at least from the 1960s to the 1990s)[1] used the slogan "airplanes don't flap their wings" to justify the claim that AI may be developed without reference to the study of biological systems. But if we reject Position (A) above–even for an AI informed by data on human performance (Newell & Simon, 1972) but not on human neurology–must we go to the other extreme of what might be called "neurochemical reductionism"?:

> POSITION (B). Any human cognitive architecture must take account of the way in which mood, emotion, and motivation affect human performance. We know that drugs can alter mood, and we know that the action of many of these drugs involves the way in which they bind to receptors in the cell membranes of neurons. Thus, no human cognitive architecture can be complete unless it incorporates the relevant specificities of neurochemistry.

Rather than discuss Position (B), my review developed an intermediate position which encourages an interchange between *distributed* AI (DAI) and cognitive neuroscience. To continue with the airplanes versus birds analogy, the bridging science of aerodynamics develops key concepts like lift, and then explains the different strategies of planes and birds in terms of the surface properties of the two kinds of wings and the way air is moved across them. In the same way, it is my hope (it is a research strategy which is yielding results but is by no means universally established) that the following approach may provide the right intermediate between Positions (A) and (B) above:

> POSITION (C). Cognitive science is to be conducted in terms of a vocabulary of interacting functional units called *schemas*. This version of *schema theory* originated as an approach to knowledge representation explicitly shaped by the need to understand how cognitive and instinctive functions can be implemented in a distributed fashion such as that involving the interaction of a multitude of brain regions or even biologically plausible neural networks. However, the functional definition of the schemas will in many cases be constrained only by the data of "brain-free" cognitive science.

[1] Minsky and Papert (1969) had published a book, *Perceptrons: An Introduction to Computational Geometry*, that many accepted as proof that artificial neural networks must fail even on tasks that were simple compared to those of then-current symbolic AI. However, the book showed only that *simple* Perceptrons (feedforward connections from inputs to one layer with adaptive synapses) suffered from this limitation (Arbib, 1969). It is thus ironic that the immense success of AI in the 21st century rests on the combination of adaptive artificial neural networks and massive and fast computing resources (LeCun et al., 2015).

The version of schema theory espoused here was an outgrowth of that offered by Arbib (1981).

(i) Schemas are ultimately defined by the execution of tasks within a physical environment. A set of *basic motor schemas* is hypothesized to provide simple, prototypical patterns of movement. These combine with *perceptual schemas* to form *assemblages* or *coordinated control programs* which interweave their activations in accordance with the current task and sensory environment to mediate more complex behaviors. Many schemas, however, may be abstracted from the perceptual-motor interface. Schema activations can be both *data-driven* (*bottom-up processing* in response to current stimuli) and *task-driven* (*top-down processing* reflecting the goals of the organism and the physical and functional requirements of the task).

(ii) A schema is both a store of knowledge and the description of a process for applying that knowledge. As such, a schema may be instantiated to form multiple schema *instances* as active copies of the process to apply that knowledge. E.g., given a schema that represents generic knowledge about some object, we may need several active instances of the schema, each suitably tuned, to subserve our perception of a different instance of the object. Schemas can become *instantiated* in response to certain patterns of input from sensory stimuli or other schema instances that are already active.

(iii) Each instance of a schema has an associated *activity level*. That of a perceptual schema represents a "confidence level" that the object represented by the schema is indeed present; while that of a motor schema may signal its "degree of readiness" to control some course of action. The activity level of a schema instance may be but one of many parameters that characterize it. Thus the perceptual schema for "ball" might include parameters to represent size, color, and velocity.

(iv) The use, representation, and recall of knowledge is mediated through the activity of a network of interacting computing agents, the *schema instances*, which between them provide processes for going from a particular situation and a particular structure of goals and tasks to a suitable course of action (which may be overt or covert, as when learning occurs without action, or the animal changes its state of readiness). This activity may involve passing of messages, changes of state (including activity level), instantiation to add new schema instances to the network, and deinstantiation to remove instances. Moreover, such activity may involve self-modification and self-organization.

(v) The key question is to understand how local schema interactions can integrate themselves to yield some overall result without explicit executive control, but rather through *cooperative computation*, a shorthand for "computation based on the competition and cooperation of concurrently active agents." For example, in **VISIONS**, a schema-based system for interpretation of visual scenes (Draper et al., 1989), schema instances represent hypotheses that particular objects occur at particular positions in a scene, so that instances may either represent conflicting hypotheses or offer mutual support. Cooperation yields a pattern of "strengthened alliances" between mutually consistent schema instances that allows them to achieve high activity levels to constitute the overall solution of a problem; competition ensures that instances which do not meet the evolving consensus lose activity, and thus are not part of this solution (though their continuing subthreshold activity may well affect later behavior). In this way, a schema network does not, in general, need a top-level executor, since schema instances can combine their effects by distributed processes of competition and cooperation, rather than the operation of an inference engine on a passive store of knowledge. This may lead to apparently emergent behavior, due to the absence of global control.

(vi) Learning is necessary because schemas are fallible. Schemas, and their connections within the schema network, must change so that over time they may well be able to handle a certain range of situations in a more adaptive way. In a general setting, there is no fixed repertoire of basic schemas. New schemas may be formed as assemblages of old schemas; but once formed a schema may be tuned by some adaptive mechanism. This tunability of schema assemblages allows them to become "primitive," much as a skill is honed into a unified whole from constituent pieces. Such tuning may be expressed at the level of schema theory itself, or may be driven by the dynamics of modification of unit interactions in some specific implementation of the schemas.

The words "brain" and "neural" do not appear in criteria (i)-(vi). I next spell out just what makes a schema-theoretical model part of brain theory:

(BTi) In brain theory, a given schema, defined functionally, may be distributed across more than one brain region; conversely, a given brain region may be involved in many schemas. Specific hypotheses about the localization of (sub)schemas in the brain may be tested by lesion experiments, with possible modification of the model (e.g., replacing one schema by several interacting schemas with different localizations) and further testing.

(BTii) Once a schema-theoretic model of some animal behavior has been refined to the point of hypotheses about the localization of schemas in various brain regions, we may further model (some of these) brain regions by seeing if its known neural circuitry can indeed be shown to implement the posited schemas. In some cases the model will involve properties of the circuitry that have not yet been tested, thus laying the ground for new experiments. In DAI, individual schemas may be implemented by artificial neural networks, or in some programming language on a "standard" (possibly distributed) computer.

Schema theory is far removed from serial symbol-based computation. However, even when this review was written (long before deep learning led to the explosion of applications in our everyday lives), work in AI had begun to contribute to schema theory, even when it did not use this term. For example, Minsky (1975) espoused a *Society of Mind* analogy in which "members of society," the agents, were analogous to schemas. Brooks (1986) developed a control scheme for robots with layers of asynchronous modules that could be considered as a version of schemas (but see Lyons & Arbib, 1989, for an explicitly schema-theoretic model of computation for sensory-based robotics). Their work shares with schema theory, with its mediation of action through a network of schemas, the point that no single, central, logical representation of the world need link perception and action–the representation of the world is *the pattern of relationships between all its partial representations*.

We may now return to the claim of Position (C) that cognitive science is to be conducted in terms of a vocabulary of interacting schemas (or schema instances), and that neuroscience may then in certain cases accept the task of explaining the properties of these schemas in terms of neural networks. Even though cognitive science itself (as distinct from AI) may be relieved of responsibility for explaining how schemas are implemented, it must still (just as a flexible feathered wing is different from a rigid metallic wing) be based, at least in part, on schemas which represent the functioning of hundreds of simultaneously active regions of the human brain. But there is nothing in the General Problem Solver (GPS) tradition initiated by Newell, Shaw, and Simon (1959), or in Newell's book, that looks at distributed processing in any detail, let alone neurological data that constrains how the different parts of the computation might be located in the different parts of the brain. The point is *not* that all good cognitive science (let alone all AI) *must be* cognitive *neuro*science. It is rather that a general framework for cognitive science must *include* cognitive neuroscience. In fact, given the current state of scientific knowledge, any current schema-level model of a cognitive system must be heterogeneous in that some schemas can be modeled in terms of detailed neural circuitry, some can be related to brain regions for which few details of circuitry are known, while others represent hypotheses about functional components for which little or no constraining neural data are available.

3 A Final Word

The original review continued for another eight pages beyond the extracts adapted above, but I will not reproduce them here–the original review is still accessible. Of course, much has happened in the 30 years since that review was first written. The chapter "From neuron to cognition: An opening perspective" (Arbib, 2016a) provides one relatively recent perspective on these developments in relation to (cognitive) neuroscience, and includes an overview of the edited volume *From Neuron to Cognition via Computational Neuroscience* (Arbib & Bonaiuto, 2016). It includes answers to "how do we get from neurons or schemas to symbols if symbols are not the building blocks of processing?" by including material on modeling brain mechanisms of language, and on pathways of biocultural evolution that endowed humans with languages (see also Arbib, 2016b).

References

Arbib, M. A. (1969). Review of 'Perceptrons: An Introduction to Computational Geometry' (Minsky, M., and Papert, S.; 1969). *IEEE Transactions on Information Theory*, *15*(6), 738-739. doi:10.1109/TIT.1969.1054388

Arbib, M. A. (1981). Perceptual structures and distributed motor control. In V. B. Brooks (ed.), *Handbook of Physiology — The Nervous System II. Motor Control* (pp. 1449-1480). Bethesda, MD: American Physiological Society.

Arbib, M. A. (1993). Book review: Allen Newell, *Unified Theories of Cognition*. *Artificial Intelligence*, *59*, 265-283.

Arbib, M. A. (2016a). From neuron to cognition: An opening perspective. In M. A. Arbib & J. J. Bonaiuto (eds.), *From neuron to cognition via computational neuroscience* (pp. 1-71). Cambridge, MA: The MIT Press.

Arbib, M. A. (2016b). Towards a computational comparative neuroprimatology: Framing the language-ready brain. *Physics of Life Reviews*, *16*, 1-54.

Arbib, M. A., & Bonaiuto, J. J. (eds.). (2016). *From neuron to cognition via computational neuroscience*. Cambridge, MA: The MIT Press.

Augusto, L. M. (2021). From symbols to knowledge systems: A. Newell and H. A. Simon's contribution to symbolic AI. *Journal of Knowledge Structures & Systems*, *2*(1), 29-62.

Brooks, R. A. (1986). A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation, RA-2*, 14-23.

Draper, B. A., Collins, R. T., Brolio, J., Hanson, A. R., & Riseman, E. M. (1989). The schema system. *International Journal of Computer Vision*, *2*, 209-250.

LeCun, Y., Bengio, Y., & Hinton, G. E. (2015). Deep learning. *Nature*, *521*(7553), 436-444. doi:10.1038/nature14539

Lyons, D. M., & Arbib, M. A. (1989). A formal model of computation for sensory-based robotics. *IEEE Transactions on Robotics and Automation*, *5*, 280-293.

Minsky, M. L. (1975). A framework for representing knowledge. In P. H. Winston (ed.), *The psychology of computer vision* (pp. 211-277). New York: McGraw-Hill.

Minsky, M. L., & Papert, S. (1969). *Perceptrons: An introduction to computational geometry*. Cambridge, MA: The MIT Press.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA & London, UK: Harvard University Press.

Newell, A., Shaw, J. C., & Simon, H. A. (1959). Report on a General Problem-Solving Program. Paper presented at the *Proceedings of the International Conference on Information Processing*. UNESCO.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Cite this article as:

Arbib, M. A. (2021). Schemas versus Symbols: A view from the 90s. *Journal of Knowledge Structures & Systems*, *2*(1), 68-74.

EDITORIAL INFORMATION

Editor-in-chief:[a] Luis M. Augusto

[a]Commentary articles are reviewed only by the editor-in-chief for appropriateness of content and style.