In Proceedings of the Sixth International Conference on Cognitive Modeling, 382-383.
Mahwah, NJ: Lawrence Earlbaum.

A General-Purpose Computational Model of the Conscious Mind

Alexei Samsonovich (asamsono@gmu.edu)
Krasnow Institute for Advanced Study, George Mason University
4400 University Dr. MS 2A1, Fairfax, VA 22030 USA

Kenneth DeJong (kdejong@gmu.edu)
Department of Computer Science and Krasnow Institute for Advanced Study
George Mason University, 4400 University Dr., Fairfax, VA 22030 USA

Elements of the Proposed Approach

Building artificial cognitive systems that possess a concept of "self", as well as cognitive and behavioral abilities of a conscious being is a difficult task, but one with tremendous practical significance and potential. In this work we outline general principles of a new approach to this task and illustrate them with detailed computational analysis based on a simple paradigm. Our approach to this task is grounded on three conceptual components – three key ideas (schemas, charts and the self) that underlie three levels of organization of our cognitive system. They are explained below.

Calculus of Schemas and Mental States

Schemas are the basic elementary building blocks of our cognitive system (Samsonovich & DeJong, 2003). The term "schema" (plural "schemata" or "schemas") was introduced by Kant (1781/1929) and is currently used in a variety of senses in computational and cognitive sciences. Here this term refers to an abstract model or a template that is used to instantiate and to process a particular cognitive category. Thus, schemas are units of semantic knowledge, primitives of action and reasoning, rules, concepts, sensory qualia, etc.

We generally say that a schema has a state, when its instance is bound to some given content. Any connected complex of states (e.g., A-D-C, Figure 1) is also regarded as a state (and is potentially convertible into a schema). For example, a state of seeing a red circle includes states of the schemas of red and a circle. A state generating a proof can be described in terms of states of schemas of logical inference, etc. Here a state is considered "mental", if it is symbolically attributed to a subject of experience (a self). Schemas and states are dynamical objects: they can be created and modified "online".

We implement schemas and their states as data structures, using one universal format, regardless of the category. Each schema has a "head" (specifying its binding terminals, rules and conditions of binding, as well as the expected effects of execution) and a "body" (specifying how, if at all, the schema is executed). Simply speaking, if a schema can be viewed as a class, then its state is an instance of this class.

Dynamical Multichart Architecture

The second key idea of our approach is the dynamical multichart architecture that serves as an infrastructure for the process of cognition. We use a simulationist approach based on an abstract notion of a chart. Each chart represents a simulated mental perspective, providing a room for a mental simulation (i.e., dynamics of mutually bound states) of what one may be expected to experience when placed in that perspective. Therefore, a chart is associated with a particular instance of the subject to whom the mental states are attributed: this is reflected in the chart label (e.g., I-Now, I-Next, I-Previous, I-Imagined, I-Goal, I-Past, I-Meta, He-Now, She-Now, etc.). Typically, this instance of the subject is the imaginary "self" of the cognitive system itself, taken in a particular context (e.g., at a particular moment of time).

Thus outlined system of charts and associated with them mental simulations can be viewed as a simplified, practical implementation of a generalized modal logic framework that may include epistemic, doxastic, deontic, conditional, temporal, alethic and other logics (with mental states regarded as "doxons", "deons", etc.). While each simulation is confined to its chart, charts interact with each other according to the rules of a corresponding logic. E.g., a state A (Figure 1) that may represent a voluntary action initiated in I-Now is copied into I-Next to produce expected effects.

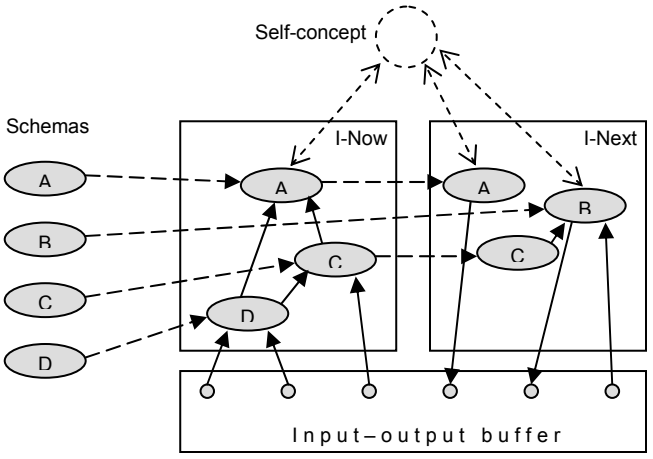

Figure 1: A snapshot of the cognitive system dynamics.

The Self Concept

The third, metacognitive component of our conceptual framework is the self concept. From the above, the reader might expect that now we shall introduce a gizmo operating on states and charts, saying that it represents the self of the

system. On the contrary, we postulate that there is no such gizmo, or virtual engine, or homunculus, etc., in any truly conscious cognitive system (we do have "drivers" operating on states and charts that follow simple, mechanistic rules).

In our approach, the self is implemented via a set of self-axioms (Aleksander & Dunmall, 2003; Samsonovich & Nadel, 2005) that globally constrain dynamics of states and charts at a higher level. In particular, this implies a linear organization of the main sequence of charts into a consistent scenario (stream of consciousness) underlying perception, thoughts, attention focus, decisions and "voluntary" actions performed by the system. This scenario (which extends in both directions in time and is continuously updated), when expressed at a verbal level (based on introspective reports that are automatically generated by states), can be controlled by linguistic rules and constraints (instantiating self-axioms) using methods of discourse analysis and the like.

Demonstration Based on a Learning Paradigm

In order to demonstrate the outlined above general principles in action, we consider a paradigm that can be called "leveraging self-learning with the self-concept". In this paradigm, a set of specially designed virtual worlds is used as a "training facility" to help a virtual robot controlled by the cognitive system to develop useful and powerful schemas. Innate schemas may include elementary moves and senses, as well as relevant reasoning primitives. The system repeats the following procedure:

1. Select an action schema and mutate its head to produce an idea of an action that is not available yet as a schema.
2. In each of several encountered situations, take the new header as a challenge and solve it (e.g., by imagined trial-and-error procedure), then execute the solution.
3. Reinterpret own behavior (not imagery) based on the schema of a voluntary action: find an apparent common motivation in the performed intermediate steps in all cases.
4. Based on the above, write the body of the new schema.

As the robot learns essential "tricks" at one level (and compiles them into schemas), it is taken to the next level. At each new stage, previously developed schemas are used for solving new challenges.

The scheme outlined above will be demonstrated in the poster by computer simulations based on a push-push puzzle setup. A minimal set of innate schemas includes a one-step move and some useful cognitive primitives, e.g., the notion of Euclidean distance. At the first stage the robot learns to move in an open space. Then it learns to navigate a maze, to push blocks, to avoid irreversible moves, etc. After that, when given a goal, it is capable of solving simple puzzle configurations and learns to deal with more complex ones.

Related Works

During recent years a tremendous progress has been made in several fields related to intelligent agents possessing higher and meta-cognitive abilities, including logical foundations (e.g., Panzarasa et al., 2002), agent architecture (e.g., works originated from the BDI framework: Bratman, 1987) and practical computational tools (e.g., Soar: Laird et al., 1987, and ACT-R: Anderson & Lebiere, 1998). The presented framework stands closer to the latter, offering more universality and a higher capacity for development and generalization than an apparatus based on chunks and/or productions. Next, our multichart architecture facilitates implementation of modal logics in a practically useful manner. Finally, our third component (the self concept) is completely missing and probably not feasible in Soar, in ACT-R, and in most other computational cognitive models.

Our system of charts can also be related to the framework of possible world boxes proposed by Nichols and Stich (2003), although each our "box" (chart) represents a possible state of agent's awareness rather than a world.

Concluding Remarks

Our present ambition is to create a virtual entity emulating human mind in its most essential abilities, an entity that by itself might be of a great scientific and practical interest for us as a general-purpose, new-generation intelligent artifact. Of a major interest for us are its abilities to learn, to exhibit rational initiative, and to be able to work interactively with a human. Among near-future applications of the proposed model or its equivalents we expect to find tasks that involve:

(a) learning to understand and to develop communication languages (e.g., learning natural language from a human),

(b) learning to do lower-level tasks based on their higher-level description (e.g., automated programming),

(c) learning to assist a human in a given cognitive domain (e.g., rapid decision making in unstable environments),

(d) self-adaptation and survival in complex environments.

References

Aleksander, I., & Dunmall, B. (2003). Necessary first-person axioms of neuroconsciousness. *Lecture Notes in Computer Science,* 2686: 630-637.

Anderson, J.R., & Lebiere, C. (1998). *The Atomic Components of Thought.* Mahwah, NJ: Erlbaum.

Bratman, M.E. (1987) *Intentions, Plans, and Practical Reason.* Harvard University Press: Cambridge, MA.

Kant, I. (1781/1929) *Critique of pure reason.* Translated by N. K. Smith. New York: St. Martin's Press.

Laird, J.E., Newell, A., & Rosenbloom, P. (1987) Soar: an architecture for general intelligence. *Artificial Intelligence* 33: 1-64.

Nichols, S., & Stich, S.P. (2003). *Mindreading: An Integrated Account of Pretense, Self-Awareness, and Understanding Other Minds.* Oxford: Clarendon.

Panzarasa, P., Jennings, N.R., & Norman, T.J. (2002) Formalizing collaborative decision-making and practical reasoning in multi-agent systems. *Journal of Logic and Computation* 12 (1): 55-117.

Samsonovich, A. V., & De Jong, K. A. (2003). Meta-cognitive architecture for team agents. *Proceedings of the 25th Annual Meeting of the Cognitive Science Society* (pp. 1029-1034). Boston, MA: Cognitive Science Society.

Samsonovich A, Nadel L (2005) Fundamental principles and mechanisms of the conscious self. *Cortex,* in press.