

Framework for Assessing AI Systems' Cognitive Architectures

Robert M. Rolfe (rolfe@ida.org) and Brian A. Haugh (bhaugh@ida.org)

Human-level artificial intelligence (AI) has many potentially beneficial applications, but research has yet to produce a computer system with the intelligence of a 5-year-old child. Under sponsorship of the Defense Advanced Research Projects Agency (DARPA), IDA developed a framework for measuring progress in the achievement of truly intelligent systems and used it to analyze existing cognitive architectures that show promise for contributing to the development of machines with human-level intelligence.

IDA's integrated cognition framework (below) is a way of organizing the essential capabilities of human cognition along the six dimensions represented by the framework's axes. The levels of cognitive capabilities typically increase as they go out the axis from the center, though some are just qualitatively different. Fully human-level cognition requires a system to have all the capabilities identified in the framework.

We used the framework to assess which of these capabilities are supported by 15 different cognitive software architectures.

The results for two of these architectures, Soar and Adaptive Control of Thought-Rational (ACT-R), are superimposed on the framework on the next page.

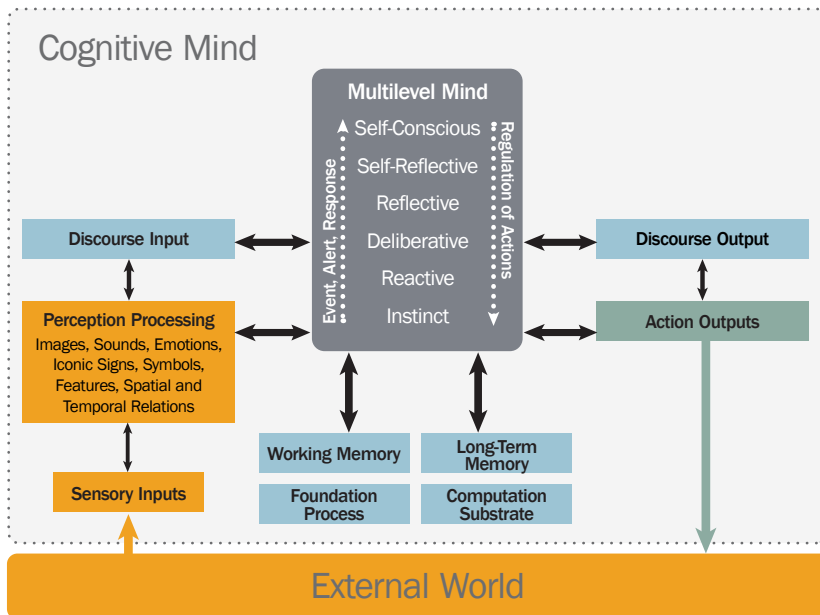
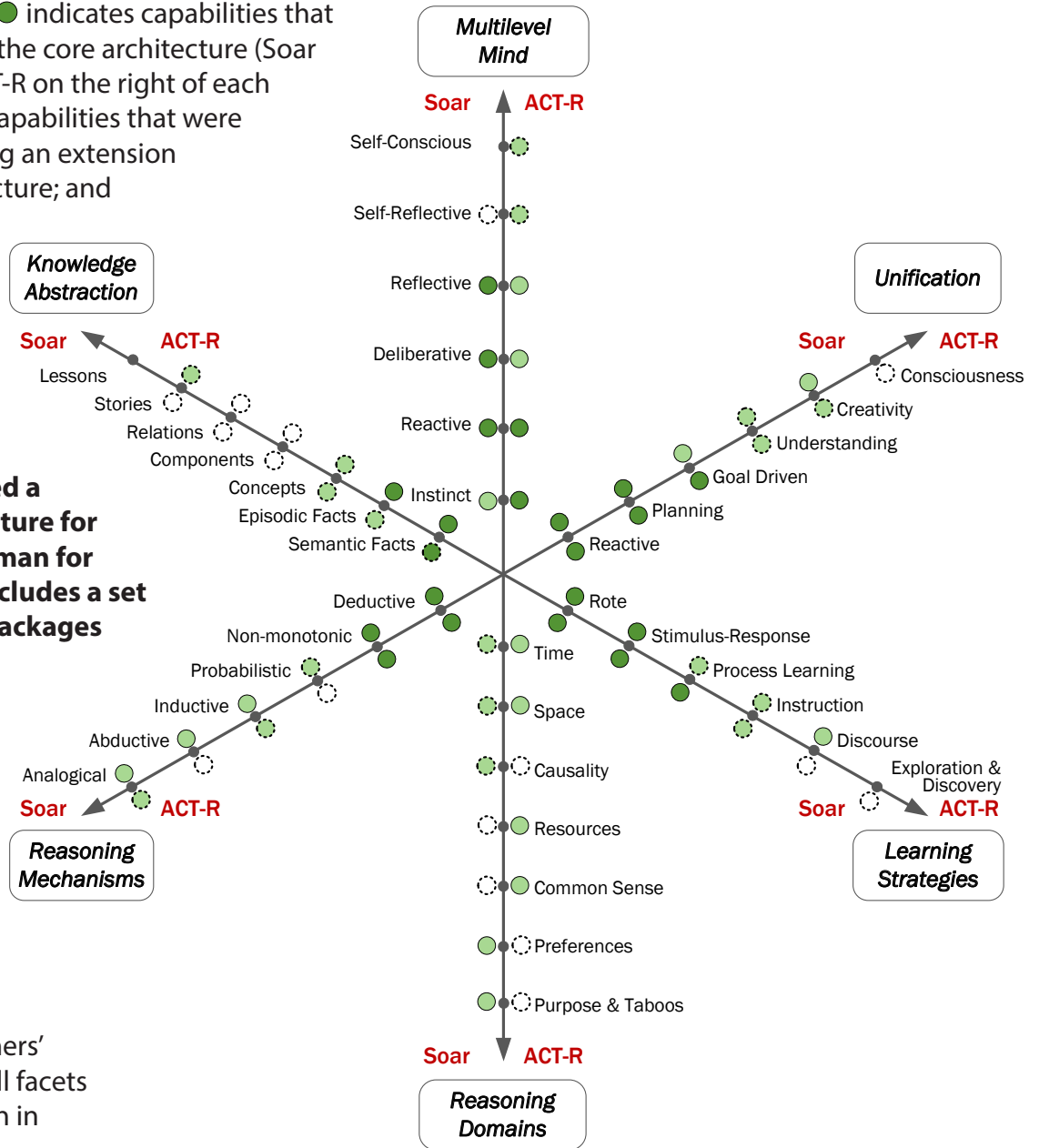


(continued)

In this framework, ● indicates capabilities that are already part of the core architecture (Soar on the left and ACT-R on the right of each axis); ○ indicates capabilities that were demonstrated using an extension to the core architecture; and ○ indicates capabilities with potential for extension of the core architecture given further investment.

IDA also developed a cognitive architecture for DARPA as a strawman for further work. It includes a set of 61 functional packages that together integrate all the facets of the cognitive dimensions identified in the framework.

Below is a top-level look at the strawman. It represents a step forward in researchers' ability to address all facets of human cognition in artificial systems.



By using new methods and tools to express the interaction of cognitive components, researchers can enable a system to understand how it operates so that it can compose itself as needed to solve specific problems, much like humans do.

Based on IDA P-3855, *Integrated Cognition—A Proposed Definition of Ingredients, a Survey of Systems, and Example Architecture*, R. M. Rolfe and B. A. Haugh, October 2004, and two papers by the same authors in AAAI Technical Report FS-17, "Integrated Cognition: A Framework Proposal," 2017, 437–442, and "Integrated Cognition: A Survey of Systems," 2017, 443–448. Research was sponsored by DARPA.