Journal of Artificial General Intelligence 4(2) 23-58, 2013
DOI: 10.2478/jagi-2013-0004

Submitted 2013-09-30
Accepted 2013-12-02

Editorial: Conceptual Commitments of AGI Systems

Haris Dindo HARIS.DINDO@UNIPA.IT
Computer Science Engineering (DICGIM)
University of Palermo
Viale delle Scienze - Edificio VI
90128 Palermo, Italy

James Marshall JMARSHALL@SARAHLAWRENCE.EDU
Computer Science Department
Sarah Lawrence College
One Mead Way
Bronxville, NY 10708, USA

Giovanni Pezzulo GIOVANNI.PEZZULO@ISTC.CNR.IT
Istituto di Scienze e Tecnologie della Cognizione (ISTC-CNR)
Via S. Martino della Battaglia, 44
00185 Roma, Italy

What are the most important design principles that we should follow to build an Artificial General Intelligence? What should be the key constituents of systems-level models of cognition and behavior?

In the target article "Conceptual Commitments of the LIDA Model of Cognition", Stan Franklin, Steve Strain, Ryan McCall, and Bernard Baars tackle these difficult problems. They propose twelve "conceptual commitments" or tentative hypotheses that form the core of the Learning Intelligent Distribution Agent (LIDA) model that they have been developing over the last ten years or so. Although the article is focused on the LIDA model, these "conceptual commitments" have much broader scope and are offered to the AGI community as specific constraints that should inform the research agenda for the realization of an Artificial General Intelligence (AGI).

The twelve specific "conceptual commitments" are of various kinds and have different degrees of importance for LIDA and AGI more generally. Some (Systems-level Modeling, Global Workspace Theory, Learning via Consciousness, Feelings as Motivators and Modulators of Learning, Transient Episodic Memory) are considered to be key for LIDA and also more broadly for AGI. These are general mechanisms of learning, memory and inference that should form the core of realistic, real-world architectures of brain and behavior. Of particular note, the authors highlight the importance (among the other things) of feeling and consciousness, which are regarded as fundamental architectural solutions to the problems of AGI. These themes, which were given minor importance in traditional cognitive (neuro)science and AI, have increasingly gained prominence in the last few years. Putting these themes at center of AGI research is a distinguishing aspect of the proposal of Franklin and collaborators.

Some other "conceptual commitments" (Biologically Inspired, Embodied or Situated Cognition, Cognitive Cycles as Cognitive Atoms, Comprehensive Decay of Representations and Memory, Asynchrony, Non-linear Dynamics Bridge to Neuroscience, Theta Gamma Coupling from the Cognitive Cycle) are very important for LIDA but not necessarily so for AGI. Most of these commitments relate to the link between AGI and neuroscience. Clearly, the question of whether or not (or to what extent) an AGI should be biologically realistic is far from settled; the authors add interesting considerations to this debate by showing the importance of biological constraints on the LIDA model.

Finally, two "conceptual commitments" (Profligacy in Learning, Consolidation) are less central to LIDA and the enterprise of AGI. Clearly, any systems-level proposal must have ancillary mechanisms that permit its functioning; deciding whether or not to elevate them to indispensable principles is again an important architectural choice, exemplified in this debate.

The target article "Conceptual Commitments of the LIDA Model of Cognition" intends to stimulate a debate in the AGI community on, first, the specific working hypotheses and design principles proposed by Franklin and co-authors; and second, and more generally, on the importance of identifying and making explicit the design principles and working hypotheses of one's own computational architecture—and even more so in the design of large scale architectures such as those targeted by the AGI community.

And indeed the target article has already generated some initial debate: the six commentaries included here have raised interesting challenges to several aspects of the proposal of Franklin and co-authors.

In their commentary, Benjamin Angerer and Stefan Schneider sound a cautionary note that AGI researchers would do well to keep in mind in developing their theories and models, namely, that the nuts-and-bolts implementation of a theory is just as important as the theoretical concepts themselves, and one must be careful to distinguish between the two. They also agree with Franklin et al. on the need to identify good general benchmarks for evaluating and comparing AGI systems, but emphasize that human cognition may have a particularly important role to play in guiding the development of these benchmarks. Finally, they point out that the integrative approach to AGI rests on a variety of implicit concepts and assumptions from cognitive psychology—assumptions that ultimately may or may not turn out to be warranted. To some extent such assumptions are necessary if we wish to build concrete models, but they nevertheless entail a certain risk. It may be necessary to rethink or abandon them to make further progress—another point of caution to keep in mind.

In his commentary, Antonio Chella applauds LIDA's commitment to consciousness as a core principle of intelligence, in contrast to other integrative cognitive architectures in which other aspects of intelligence such as problem solving, resource maximization, or integration of capabilities are regarded as the key principles. He suggests that this focus on consciousness is itself an important conceptual commitment for AGI that should be included in the list of commitments proposed by Franklin et al.

In his commentary, John Laird highlights some of the differences between the SOAR cognitive architecture and LIDA, particularly the relative emphasis that each architecture places on functional commitments versus biological or psychological commitments. He outlines a set of general functional constraints and requirements for AGI systems, and emphasizes the importance of real-time system performance. In his view, LIDA's attempt to incorporate functional, biological, and psychological constraints within a single system may be overly ambitious, at least if efficient real-time performance of the system is also a requirement of the model.

Olivier Georgeon and David Aha focus on Franklin et al.'s conceptual commitment "Cognitive Cycles as Cognitive Atoms". This commitment is central to the LIDA model, but Franklin et al. are undecided as to its level of importance for AGI systems in general. Georgeon and Aha, however, view it as critically important, and propose an even stronger conceptual commitment called "Radical Interactionism" (RI), which recasts the basic principle of an indivisible cognitive cycle (a "cognitive atom") in terms of sensorimotor interactions. In their view, the traditional distinction between perception and action as separate entities is unnecessary and misleading. Instead, their RI commitment subsumes perception and action into the more fundamental notion of sensorimotor interaction, which they consider to be the appropriate primitive on which to base cognitive agent architectures. They also show how the LIDA architecture could be modified to reflect this new conceptual interpretation.

In his commentary, Pei Wang reflects thoughtfully on the unique challenges faced by the field of AGI in designing general-purpose AI systems, due in part to its lack of established, agreed-upon theories and frameworks to guide research. He points out that the conceptual commitments that underlie different AGI projects may differ according to which aspects of human intelligence the projects focus on. That is, the particular research objectives of a project may determine which conceptual commitments are relevant, and these commitments may overlap only partially with those proposed by Franklin et al. He also considers the different types of challenges and risks that arise in taking an integrated versus a unified approach to AGI, and makes the important point that being able to describe some aspect of intelligence in psychological terms is not by itself sufficient justification for its implementation in an AGI system as a distinct module or mechanism.

In their commentary, Travis Wiltshire and his colleagues take issue with Franklin et al.'s apparent commitment to "disembodied embodiment". They raise important questions regarding the LIDA architecture's level of commitment to feedback-rich agent-environment interaction in general, as well as to socially interactive capabilities in particular. More broadly, they suggest that a stronger commitment to human-like embodiment (as opposed to Franklin et al.'s less specific notion) may be necessary in order to achieve AGI's ultimate goals. Furthermore, they point out that in evaluating an AGI system, it is crucial to take into account the extent to which the system is perceived and treated by humans as conveying agency through social interactivity.

In summary, this issue of the JAGI journal hosts a stimulating debate on which design principles and "conceptual commitments" should form the foundations of large-scale systems for Artificial General Intelligence. Stan Franklin, Steve Strain, Ryan McCall, and Bernard Baars propose a rich set of important working principles that stem from their long experience with their Learning Intelligent Distribution Agent (LIDA) model. The specific principles are a matter of debate: they can be discussed, elaborated on, and called into question—and indeed the lively discussion provided by the commentators testifies that this debate has already begun. Still, the authors are, in our opinion, correct in highlighting that the time has come to distill key principles from current research in cognitive science, neuroscience, AI, machine learning, and beyond, that these principles need to be operationalized and made explicit, and that their discussion is a key research objective of the AGI (and JAGI) community. We sincerely hope that this JAGI special issue will contribute significantly to this important objective, and will stimulate the type of long-lasting debate that is crucial for the overall progress of the discipline.