A Formal Analysis of Intelligent Agents with Mathematical Tools

Zippora Arzi-Gonczarowski (zippie@actcom.co.il)
Typographics, Ltd.; 46 Hehalutz Street
Jerusalem 96222, Israel

ISAAC ('Integrated Schema for Affective Artificial Cognition') is a mathematical model of intelligent agents, which gives rise to a formal theory that could be implemented computationally.

Similar to the natural evolutionary context, the schema starts from a simple model of corresponding sensations and reactions. It then structures 'upgrades' (e.g. handle conflicting reactions, internal representation, and so on) on top of that, using generative reasoning to systematically obtain and study the properties of these upgraded structures. Among other things, this approach yields a continuous bridge from low level to high level intelligence.

A *perception* snapshot is structured as a set of *world elements* that constitutes an environment (real or imagined), a set of *connotations* that constitutes a collection of discriminations, and a set of *behaviors*. Behaviors are conjured up on the basis of a *perception predicate* that relates between world elements and their connotations in a three valued manner (*true, false, undefined*). With real environments, that basic schema approximates models of simple forms of intelligence. For higher level forms of intelligence, mind activities (cognitive, behavioral, affective) are modeled as streams of perceptions. Along these streams, all the components mentioned above could adapt dynamically: be modified, extended, merged, and so on. In mathematical terminology, perceptions are *objects*, and passages from one perception to another are modeled as *morphisms*.

In the course of a few years of ongoing research, a variety of mind processes have been modeled on the basis of these uniform, yet flexible, premises, capturing mental activities from streams of interpretations, through behavior development and integration, representation formation, imaginative design and anticipation, analogy making, to social and self perception. Each publication has been appraised as a model of the relevant mind aspect, but as a collection they also feature an additional value of an integrated whole: because they share uniform modeling premises, the various processes can be neatly composed and alternated between, modeling multifaceted intelligences.

Mathematical treatments are naturally expected to come up with equations. The equational results of ISAAC are categorical commutative diagrams that state where one formalized mind process is a 'paraphrase' of another. These diagrams are like high level 'blue prints' that provide a computational schema with basic structural guidelines for neat agent architectures.

Mathematization has, indeed, proven a powerful modeling tool for other scientific domains. In cognitive science, however, the open ended diversity of phenomena that need to be modeled has made it difficult to capture things in a uniform manner. If one were to overcome that obstacle, then this would probably be by finding a suitable level of abstraction: high enough to absorb a variety of phenomena, but not too high so that meaningful things could still be stated. Mathematical categorization is a tool that has been developed precisely for such purposes within mathematics itself. ISAAC deploys well developed category theoretical tools for cognitive science purposes, yielding a general, yet rigorous, schema.

Pre-theoretically, ISAAC is grounded by intuitions. As a theory, the treatment proceeds as if the semantic primitives were context independent: All definitions, constructions and results are tidily operated within the abstract mathematical framework. Then, they are invariably examined with regard to the grounding intuitions, pre-theoretical conceptions, and existing knowledge about the modeled processes. Results that have not been anticipated at the outset provide supporting arguments that the proposal is apparently on a promising track.

Substitution instances of the schema could be applied to describe, or to design, particular agents, filling in detailed domain-specific features to approximate particular agents. The foundational schema avoids over determinism and captures abstracted structures and mechanisms that are shared by intelligent agents. This yields a unifying theory and, hopefully, contributes to:
(1) Cognitive science becoming a science, and to
(2) Integrative artificial intelligence that does not lose the big picture by over fragmentation.

References

The publications can be found on the author's web page: `www.actcom.co.il/typographics/zippie`