

Review of *Metareasoning: Thinking about thinking*, ed. Michael Cox and Anita Raja, MIT Press, 2011, vii+340 pages.

Metareasoning is reasoning about a reasoning process. This book is a collection of twenty recent research papers by different authors, stemming from a workshop in Chicago in July 2008.

The term “metareasoning” has been applied to many different ideas and techniques. In order to “assemble some measure of consistency and soundness in the topic”, the editors wrote a short “manifesto” laying out a very general, simple model of metareasoning. The authors relate their own work to this framework; they also discuss the relation of their work to the other papers in the collection.

The manifesto discusses three aspects of metareasoning. First, there is a three-level model of an agent: at the ground level, the agent acts in an environment; at the object level, it reasons about its actions and the environment; at the metalevel, it reasons about its ground level reasoning. For instance, if the agent is using an anytime algorithm to generate a plan, then it must eventually decide that the current plan is acceptable and that it should act now rather than spend more time thinking; this decision is made at the metalevel. Second, there is multi-agent metareasoning; in a group of interacting agents, each agent reasons about the reasoning processes of the rest. Third, there is self-modeling: the agent’s model of the world should include itself, and in particular should include knowledge of its own reasoning abilities. The papers in fact mostly lie close to this framework.

I found this book unsatisfying, at two levels. First, in many of the papers, the description of the object level is inadequate, perhaps because of the 16 page limit. For example, “Using introspective reasoning to improve a CBR system performance,” by Arcos, Mülâyim, and Leake describes a case-based reasoning system used in the design of gas treatment plants. The application and approach sound interesting, but there is no description of what the system does or how it works. One or two examples discussing a specific problem, the comparable cases that were invoked, and the role of metareasoning in picking good cases, would be much more meaningful than the result that four different metareasoning strategies reduced the fraction of “Low-Quality” cases retrieved from 16.67% to 14.49%, 9.05%, 12.85%, and 8.21%. Many of the other papers have comparable gaps.

The second source of dissatisfaction is that there is little attempt to situate the research directions discussed here within the broader context of AI and CS. For instance, several papers deal with some form of scheduling, a subject with an immense CS literature that never invokes metareasoning. Quite possibly, the particular problems here have unusual features that make metareasoning more effective than conventional scheduling techniques, but there is no discussion of what those features are.

Another example: Kennedy in “Distributed Metamanagement” writes, “[A] robot expects to find an office at the end of a corridor, but instead finds a cupboard. Without metalevel reasoning ... it would just assume that the cupboard is a very small office.” But this is just ordinary belief update; why is metareasoning more appropriate here than Bayesian updating, belief revision, or nonmonotonic inference?

The same issue can be viewed from the opposite direction. In his influential paper “Algorithm = Logic + Control” (*CACM*, 1979) Kowalski argued that an effective way of modularizing algorithm design is to view the computation as ground inference guided by a control regime. From this standpoint, some substantial fraction of CS — dynamic programming, compiler optimization, database query optimization, adaptive mesh techniques, etc. — can be viewed as manipulating the control regime to get the most effective inference. Closer to the areas discussed in this book, the selective deepening technique, used with great success in Deep Thought, of choosing particular nodes in a game tree to explore more deeply, conforms precisely to Cox and Raja’s schema: The ground level is the actual execution of chess moves, the object level is the evaluation of the game tree, and the metalevel is the selective deepening decision. None of these is generally described as metareasoning.

Should they be? If not, what are the features that divide them from the work described in Cox and Raja?