Luger, G.F. (2012). Can artificial intelligence improve human reasoning? A review in *IEEE Computing in Science and Engineering July/August* of R. Kowalski, (2011) Computational logic and human thinking: How to be artificially intelligent. Cambridge UK: University Press.

Can Artificial Intelligence Improve Human Reasoning?

By George F. Luger

Robert Kowalski, *Computational Logic and Human Thinking: How to Be Artificially Intelligent*, Cambridge University Press, 2011, ISBN: 9780521123365, 310 pp.

or readers desiring to know how the rational human mind might be understood and enhanced through computer-based representations and algorithms, *Computational Logic and Human Thinking* is a must-read. The subtitle, *How to Be Artificially Intelligent*, offers a further challenge that suggests that this book might also make an important contribution to understanding how humans think, justify their thoughts, and present coherent arguments.

Kowalski's goal is to "attempt to show that the practical benefits of computational logic are not limited to Mathematics and Artificial Intelligence, but can also be enjoyed by ordinary people in everyday life, without the use of mathematical notation" (p. 2). Logic, Kowalski suggests, in the tradition of Aristotle and Boole, offers a formalization for human laws of thought. Aristotle developed his theory of logic and reasoning in the *Organon*, or *The Instrument*, a term used by commentators to collect his various writings on logical forms. But Aristotle's logical forms also appeared in his *Rhetoric*—for example, I–II, where he suggested that their use is valuable for developing clear thought processes and presenting convincing arguments. Kowalski, following Aristotle, has also made computational logic the unifying theme for developing critical human skills for thinking and reasoning: logic "focuses on the formulation of normative theories, which prescribe how people ought to think. Viewed in this way, Computational Logic can be regarded as a formalization of the language of human thought" (p. 2).

The Details

The book is written in an easy-flowing and example-driven style. Chapter 1, for example, takes the set of instructions that make up an "emergency notice" on the London underground subway system and translates these into the conditional language of logic, including the *if-then*, *and*, *or*, and *not* connectives. In Chapter 2, Kowalski addresses arguments proposed by cognitive psychologists suggesting that humans aren't logic-based reasoning systems. In the following chapters, Kowalski uses another example—Aesop's fable of the fox and the crow—to introduce the notions of backward, or goal-driven, and forward, or data-driven, reasoning. Using these different reasoning modes, the human agent produces a search process, moving across a set of possible world states in the process of trying to build an argument or accomplish a task.

Because the reasoning scheme (or controlling process) of choice for Kowalski's computational logic is a form of resolution, Chapter 5 makes an argument for the use of *negation as failure* in the context of human reasoning and decision making—so basically, the failure to find a fact true can lead to the justification of it being false. This is a reasonable choice, I feel, when an agent has knowledge of all the relevant facts (the *closed-world* assumption). However, the argument is unconvincing when, as usually happens in human reasoning, such an assumption isn't warranted.

In the following chapters, Kowalski proposes a form of Emil Post's[1] and Allen Newell and Herbert Simon's[2] traditional production system as a cognitive controller, or architecture, that is able to apply goal- and data-driven reasoning schemes. This software architecture has been shown by cognitive psychologists to support many of the observable features of human problem-solving performance. As part of the decision-making component of production systems—the so-called *conflict-resolution* scheme—Kowalski relates his approach to the modern decision theory of Daniel Kahneman and Shane Frederick.[3] Kowalski invokes the prisoners' dilemma problem (the classic conflict between individual self-interests and collective interests in various contexts), including strategies and possible solutions, to demonstrate the addition of decision theory to conflict resolution to control the production system architecture.

To complete his presentation, Kowalski describes other related aspects of human decision making including the meaning of life, abductive inference,[4] agents in an evolving world of goals and purposes, and metalogic. In the book's final component, Kowalski—addressing related arguments from the psychological community—argues for the cognitive plausibility of his computational logic formalization of human reasoning and decision making.

Analysis

Kowalski has been careful, in his goal of making the book accessible to the general interested reader, to exclude from his writing almost all mathematical notation. To complement this less formal presentation in the book's body, he has included six appendices that offer a rigorous formalization that supports his earlier, more intuitive presentation of the chapters. These appendices serve as excellent tutorials for the motivated reader. Their content includes the syntax of logical forms; logic-based truth as Herbrand interpretations; forward and backward reasoning schemes (including claims for soundness and completeness); minimal models and negation; resolution refutation systems; and abductive logic programming.

Although Kowalski has kept equations out of the 17 chapters of his primary presentation, his writing isn't free from artificial intelligence and computer science technical jargon. He uses such science-specific terms as closed-world assumption, minimum models, meta-logic, default or defeasible reasoning, negation as failure, compile/decompile, encapsulation, and many more. The author could have done a great service to the general reader by avoiding these terms whenever possible; or when such terminology was necessary, by describing their meanings in a way that's easily understandable to a nonspecialist. In future editions, it would be helpful to provide a glossary of such terms to assist some readers.

What I liked best about this book is that, as his top-level goal, Kowalski has taken on the task of developing a full epistemological stance. This stance attempts to address, in the form of a computational model with related search algorithms, what it means for the human agent to perceive stimuli, reason about their meaning, and respond appropriately within the constraints of an ever-evolving world. Kowalski's enterprise is much broader than piecemeal research on issues such as goal-reduction algorithms or truth in minimal models.

The cognitive architecture Kowalski chooses for his epistemological stance—the production system—has a long history in cognitive science, with its earliest use and justification by Turing Award winners Newell and Simon.[2] In more recent years, it has been extended, in the Soar architecture, to include agent learning of new rules/skills.[5] The broad epistemological viewpoint developed throughout this book, with its goals of understanding the nature of human reasoning and improving its use in the challenges of normal life, reflect many of the insights Kowalski has garnered during more than 40 years as a highly successful researcher in the fields of artificial intelligence and computational logic.

In this book, Kowlaski takes on an enormous task—which, of necessity, requires assessing the nature, representations, and processes enabling human reasoning. So it's only natural that there will be a number of epistemological issues that remain unresolved. For example, Kowlaski's adoption of the production system architecture of the 1970s and 1980s as the brain's "software" that controls decision making,[2] is seen today as an example of Daniel Dennett's *Cartesian theater*,[6] a remnant of Descartes' dualism. This theater is a hypothesized special place in the cortex where decision theoretic algorithms sort through specific choices brought forward by sensory, emotional, memory-based, linguistic, and other components of the human agent. Modern science proposes a much more decentralized and distributed architecture for cognition, a *society of mind*[7] with key constituent roles played by multiple distributed elements of the human system. For example, the finger moves from the hot stove more quickly than any nerve signal can go from the fingertip to the brain's Cartesian theater and from there to a process for motor control and finger withdrawal.

A further concern is how the components of Kowalski's "thought processes" are to be reified (made concrete) in the form of logic expressions, as he proposes in Chapter 2. Although logic is both a convenient and suitably expressive representation, when it's then coupled with specific search algorithms and related assumptions (that is, a closed world), the entire system becomes impossible to establish as a necessary model for human reasoning in any scientific sense. This confirmation problem is called *representational indeterminacy* by the psychologist John Anderson[8] and other philosophers of science.

In this serious and enjoyable book, Kowalski proposes a specific, utilitarian, and sufficient model, in the scientific sense, of human subject/world communications. And, as Aristotle suggested long ago, the sufficiency of this logic-based representational effort could offer insights that can lead to more coherent reasoning, writing, discussions, and arguments by human agents.

References

1. E. Post, "Formal Reductions of the General Combinatorial Decision Problem," *Am. J. Mathematics*, vol. 65, no. 2, 1943, pp. 197–215.
2. A. Newell and H.A. Simon, *Human Problem Solving*, Prentice Hall, 1972.
3. D. Kahneman and S. Frederick, "Representativeness Revisited: Attribute Substitution in Intuitive Judgment," *Heuristics and Intuitive Judgment: Extensions and Application*, Cambridge Univ. Press, 2002.
4. C.S. Peirce, *Collected Papers 1931–1958*, Harvard Univ. Press, 1958.
5. A. Newell, *Unified Theories of Cognition*, Harvard Univ. Press, 1990.
6. D.C. Dennett, *Consciousness Explained*, Little Brown, 1991.
7. M. Minsky, *The Society of Mind*, Simon and Schuster, 1985.
8. J.R. Anderson, "Arguments Concerning Representations for Mental Imagery," *Psychological Rev.*, vol. 84, no. 4, 1978, pp. 249–277.

George Luger is a professor in the Computer Science, Psychology, and Linguistics Departments at the University of New Mexico in Albuquerque. His current research focuses on the development of probabilistic models of human skilled performance. Luger has a PhD in interdisciplinary research that included computer science, mathematics, linguistics, and psychology from the University of Pennsylvania. His book *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*, first published in 1989, is now in its sixth edition; his book *Cognitive Science: The Science of Intelligent Systems* was published in 1995. Contact him at luger@cs.unm.edu.