Can Artificial Intelligence Improve Human Reasoning

A Review by

George F. Luger
Professor of Computer Science Linguistics and Psychology
University of New Mexico

of

Computational Logic and Human Thinking: How to be Artificially Intelligent
By Robert Kowalski
Cambridge University Press, 2011, 310 pages

For readers desiring to know how the rational human mind might be understood and enhanced through computer-based representations and algorithms this title 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". Logic, suggests Kowalski, 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, eg., 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."

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 are not 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 backwards, or goal-driven, and forwards, 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—i.e., the failure to find a fact true can lead to the justification of it being false. This is a reasonable choice when an agent has knowledge of all relevant facts (the "closed world" assumption). This argument is unconvincing when, as usually happens in human reasoning, such an assumption is not warranted.

In the following chapters Kowalski proposes a form of the traditional production system of Post and Newell and Simon (1972) as a cognitive controller, or architecture, that is able to apply the goal- and data-driven reasoning schemes proposed. 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 Kahneman and Frederick (2002). The prisoners' dilemma problem, including its strategies and possible solutions, demonstrates the addition of decision theory to conflict resolution to control the production system.

To complete his presentation, Kowalski describes other related aspects of human decision making including: the "meaning of life", abductive inference (Peirce 1958), agents in an evolving world of goals and purposes, and meta-logic. In this final component of the book Kowalski, addressing related arguments from the psychological community, argues for the cognitive plausibility of his computational logic formalization of human reasoning and decision-making.

Kowalski has been careful, in his goal of making his book accessible to the general interested reader, to exclude from his writing almost all mathematical notation. To complement this less formal presentation in the body of the book, he has included six appendices that offer a rigorous formalization that supports his earlier more intuitive presentation of the earlier 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 seventeen chapters of his primary presentation, his writing is not free from Artificial Intelligence and Computer Science technical vocabulary. These science specific terms include *closed world assumption*, *minimum models*, *meta-logic*, *default* or *defeasible* reasoning, *negation as failure*, *compile/decompile*, *encapsulation*, and many more. The author would have done the general reader a great service by avoiding these terms when possible, or when their use is necessary, describing their meanings in terms understandable by the non-specialist. This reviewer suggests creating in future editions a glossary of such terms to assist the reader not acquainted with these terms.

What I liked best about this book is that, as his top-level goal, Robert 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 like 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 (1972). In more recent years it has been extended, in the SOAR architecture, to include agent learning of new rules/skills (Newell 1990). 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 his forty plus years as a very successful researcher in the fields of Artificial Intelligence and Computational Logic.

In Computational Logic and Human Thinking, Kowlaski takes on an enormous task, which, of necessity, requires assessing the nature, representations, and processes enabling human reasoning. It is only natural that there be a number of epistemological issues that remain unresolved. For example, Kowlaski's adoption of the production system architecture of the 1970-80s (Newell and Simon 1972), as the brain's "software" that controls decision making, is seen today as an example of Dennett's (1991) *Cartesian theater*, a remnant of Descartes' dualism. This theater is an hypothesized special place in 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 (Minsky 1985) 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 is proposed in Chapter 2. Although logic is both a convenient and suitably expressive representation, when it is then coupled with specific search algorithms and related assumptions, e.g., 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 (1978) and other philosophers of science.

Robert Kowalski, in this serious and enjoyable book, proposes a very 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 may well offer insights that can lead to more coherent reasoning, writing, discussions, and arguments by human agents.

After receiving his PhD at the University of Pennsylvania in 1973, George Luger was a post-doctoral Research Fellow at the Department of Artificial Intelligence at the University of Edinburgh, working there with Bob Kowalski in 1974 and 1975. In 1979, George went to the University of New Mexico in Albuquerque, where he is currently a Professor in the Computer Science, Psychology, and Linguistics Departments. His book *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*, first published in 1989, is now in its 6th edition. His book, *Cognitive Science: The Science of Intelligent Systems* was published in 1995. His current research focuses on the development of probabilistic models of human skilled performance.

References:

Anderson, J.R., 1978. Arguments concerning representations for mental imagery. *Psychological Review*, 84 249-277.

Dennett, D.C., 1991. *Consciousness Explained*. Boston: Little Brown.

Kahneman, D. and Frederick, S. 2002. Representativeness revisited: attribute substitution in intuitive judgment. In Heuristics and Intuitive Judgment: Extensions and Application, T. Gilovich, D. Griffen, and D. Kahnamen (eds). New York: Cambridge University Press.

Minsky, M, 1985. *The Society of Mind*. New York: Simon and Schuster.

Newell, A. 1990. *Unified Theories of Cognition.* Cambridge, MA: Harvard University Press.

Newell, A. and Simon, H.A., 1972. *Human Problem Solving.* Englewood Cliffs NJ: Prentice Hall.

Peirce, C.S., 1958. *Collected Papers 1931-1958*. Cambridge MA: Harvard University Press.