Book review

by **Luís Moniz Pereira**

http://centria.di.fct.unl.pt/~lmp/
Centro de Inteligência Artificial (CENTRIA)
Faculdade de Ciências e Tecnologia
Universidade Nova de Lisboa, Portugal

Computational Logic and Human Thinking: How to be Artificially Intelligent
by **Robert Kowalski**
Cambridge University Press, 2011
ISBN: 978-0521123365, paperback, $46.00
xxii+310 pages

This is a review of the recent book by Robert Kowalski, Emeritus Professor in the Department of Computing Imperial College, London, http://www.doc.ic.ac.uk/~rak/
In 2011 he received the IJCAI Award for Research Excellence for his contributions to logic for knowledge representation and problem solving, including his pioneering work on automated theorem proving and logic programming.

Product Description (from C.U.P)

The practical benefits of computational logic need not be limited to mathematics and computing. As this book shows, ordinary people in their everyday lives can profit from the recent advances that have been developed for artificial intelligence. The book draws upon related developments in various fields from philosophy to psychology and law. It pays special attention to the integration of logic with decision theory, and the use of logic to improve the clarity and coherence of communication in natural languages such as English. This book is essential reading for teachers and researchers who may be out of touch with the latest developments in computational logic. It will also be useful in any undergraduate course that teaches practical thinking, problem solving or communication skills. Its informal presentation makes the book accessible to readers from any background, but optional, more formal, chapters are also included for those who are more technically oriented.

Detailed Summary of the Book by Chapters (abridged from the book)

Introduction. In Artificial Intelligence (AI), an *agent* is any entity, embedded in a real or artificial world, that can observe the changing world and perform actions on the world to maintain itself in a harmonious relationship with the world. Computational Logic (CL), as used in AI, is the agent's language of thought. Sentences expressed in this language represent the agent's beliefs about the world as it is and its goals for the way it would like it to be. The agent uses its goals and beliefs to control its behaviour. It can also use CL to guide its public communications with other agents.

Logic on the Underground. The London Underground Emergency Notice illustrates the way in which the meanings of English communications can be understood as

thoughts in logical form. In CL these thoughts have both a logical and computational character. Because of this dual character, sentences expressed in this form are also called *logic programs*.

The Psychology of Logic. The most influential and widely cited argument against logic comes from psychological experiments about reasoning with natural language sentences in conditional form. The most popular interpretation of these experiments is that people do not have a natural general-purpose ability to reason logically, but have developed instead, through the mechanisms of Darwinian evolution, specialised algorithms for solving typical problems that arise in their environment. This chapter argues that one of the problems with the experiments is that they fail to appreciate that the natural language form of a conditional is only an approximation to the logical form of its intended meaning. Another problem is that the interpretation of these experiments is based upon an inadequate understanding of the relationship between knowledge and reasoning In contrast, the CL understanding of human thinking can be expressed loosely as: *thinking = specialised knowledge + general-purpose reasoning.*

The Fox and the Crow. Aesop's fable of the fox and crow illustrates the backward reasoning of a clever fox, to generate a plan to achieve the goal of having the cheese of a not so clever crow. The chapter contrasts the fox's *proactive*, backward reasoning with the crow's *reactive*, forward reasoning, to respond to the fox's praise by breaking out in song, thereby dropping the cheese to the ground, where the fox can pick it up. Both fox and crow reason according to the general-purpose inference rules of CL, but the fox has a better knowledge of the world, and more powerful ways of using it for her benefit. If the crow knew as much as the fox and were able to reason *preactively*, thinking before acting, then he could reason forward from the hypothetical performance of his candidate actions, predict their likely consequences, and choose an alternative action to achieve a better expected resulting state of affairs.

Search. In CL, a *proof procedure* consists of a collection of inference rules and a search strategy. The inference rules determine both the structure of proofs and the *search space* of all possible proofs relevant to the solution of a goal. The *search strategy* determines the manner in which the search space is explored in the search for a solution. Many different search strategies are possible.

Negation as Failure. In the semantics of CL the world is a positive place, which can be characterised by the positive atomic sentences that are true at the time. Because the ultimate purpose of an agent's goals and beliefs is to manage its interactions with the world, the syntactic form of the agent's thoughts also has a corresponding positive bias. Syntactically negative thoughts commonly arise from the *failure* to observe or derive positive information. Negation as failure is a natural way to reason by *default* with incomplete information, deriving conclusions under the assumption that the agent knows it all, but then withdrawing the conclusions if new information shows they do not hold. It also facilitates organising goals and beliefs into hierarchies of rules and exceptions.

How to Become a British Citizen. The British Nationality Act is a body of English sentences, which states precisely the conditions under which a person may acquire, renounce or be deprived of British citizenship. The Act is designed to be both unambiguous, so there is little doubt about its intended meaning, and flexible, so that

it can be applied to changing circumstances. Its English style resembles the conditional form of sentences in CL. In addition to its use of conditional form, the Act illustrates many other important features of CL, such as the representation of rules and exceptions, and meta-level reasoning about what it takes for a person to satisfy the requirements for naturalisation as a British citizen. In contrast, the University of Michigan Lease Termination Clause shows how an ambiguous, virtually unintelligible English text can be made understandable by reformulating it in CL.

The Louse and the Mars Explorer. The most influential computational model of human thinking in Cognitive Psychology is the Production Systems, model illustrated in this chapter by the wood louse and the Mars explorer robot. Production systems combine a *working memory* of atomic facts with *condition-action rules* of the form *if conditions then actions*. The working memory is like a model of the current state of the world, and the rules like an agent's goals and beliefs. Condition-action rules are embedded in an *observation-thought-decision-action cycle* and executed by matching the conditions of rules with facts in the working memory, thereby generating the actions of rules as candidate actions by *forward chaining*, which is similar to forward reasoning. If more than one candidate action is generated, *conflict resolution* decides amongst them. The chosen action is then executed, changing the state of the working memory, simulating the way an agent's actions change the state of the world. From a logical point of view, there are three kinds of condition-action rules: *reactive rules*, which are like instinctive stimulus-response associations; *goal-reduction rules*, which reduce goals to subgoals by forward chaining; and *forward reasoning rules*, which perform genuine logical forward reasoning.

Maintenance Goals as the Driving Force of Life. The agent model in the book combines the functionalities of logic and production systems in a logical framework. It takes from production systems the *observation-thought-decision-action cycle*, but replaces condition-action rules by goals and beliefs in the logical form of conditionals. It replaces reactive rules by *maintenance goals* used to reason forwards, goal-reduction rules by *beliefs* used to reason backwards, and forward reasoning rules by *beliefs* used to reason forwards. In the logical agent model, the agent *cycle* responds to observations of the environment by reasoning forwards with beliefs, until it derives a conclusion that matches one of the conditions of a maintenance goal. It reasons backwards, to check the other conditions of the maintenance goal. If all the conditions of the maintenance goal are shown to hold, it reasons forwards one step, deriving the conclusion of the maintenance goal as an *achievement goal*. It then reasons backwards using its beliefs to reduce the achievement goal to a plan of candidate actions.

The Meaning of Life. The logical framework of the book views an agent's life as controlled by the changes taking place in the world, by its own goals and beliefs, and by the choices the agent makes between different ways of achieving its goals. The combination of its beliefs and its highest-level goals generates a hierarchy of goals and subgoals. For the sake of efficiency, this hierarchy may be collapsed into a collection of more direct stimulus-response associations, whose original goals are no longer apparent, but implicit and emergent. In AI, and Computing more generally, it is common for an intelligent designer to implement an artificial agent that does not contain an explicit representation of its higher-level goals. The designer is aware of the agent's goals, but the agent itself is not. As far as the agent is concerned, its life

seems entirely meaningless. In this chapter the seemingly meaningless life of an imaginary, artificial wood louse, is compared with the more meaningful life of an intelligent agent, in which stimulus-response associations and awareness of higher-level goals are combined.

Abduction. One of the main functions of an agent's beliefs is to represent causal relationships between its experiences. The agent uses these causal representations both *proactively* to generate plans to achieve its goals, and *preactively* to derive consequences of candidate actions to help it choose between alternative candidate actions. The agent can use the same causal beliefs *abductively*, to generate hypotheses to explain its observations, and deductively to derive consequences of hypotheses to help it choose between alternatives. The process of generating and choosing hypotheses to explain observations is called *abduction*. Like default reasoning with negation as failure, abduction is *defeasible*.

The Prisoner's Dilemma. Deciding between alternative abductive explanations of an observation is similar to deciding between alternative actions, exemplified by the Prisoner's Dilemma. In this chapter, it is seen how an agent can use a combination of CL and decision theory to choose between alternatives. In decision theory the agent should choose an alternative having the best expected outcome, determined by combining judgements of the utility of the actions' consequences with judgements of the likelihood that their consequences will actually happen. Decision theory is normative, requiring detailed knowledge of utilities and probabilities, but neglecting the motivations of an agent's actions. In practice, agents typically employ heuristic goals and beliefs, to approximate the decision-theoretic norms. To make smarter choices than those obtained by decision theory or heuristics alone, it is better to use the broader framework of the agent cycle, to analyse the motivations of actions, and to ensure that a full range of alternatives is explored.

Motivations Matter. Decision Theory leads to consequentialist theories of morality, which judge the moral status of actions simply in terms of their consequences. But in psychological studies and the law, people judge actions both in terms of their consequences and their motivations. CL can model such moral judgements by using constraints to prevent actions deemed morally or legally unacceptable.

The Changing World. An agent's life is a continuous struggle to maintain a harmonious relationship with the ever-changing world. The agent assimilates its observations of the changing state of the world, and performs actions to change the world in return. To help it survive and prosper in such a changing environment, an agent must use beliefs about cause and effect, represented in its language of thought. This chapter investigates the logical representation and semantics of such causal beliefs.

Logic and Objects. Whereas in Cognitive Psychology production systems are the main competitor of Logic, in Computing the main competitor is Object-Orientation (OO). In the OO point of view, the world consists of objects that interact by sending and receiving messages. Objects respond to messages by using encapsulated methods, invisible to other objects, and inherited from methods associated with general classes of objects. CL is compatible with OO, if objects are viewed as agents, methods are viewed as goals and beliefs, and messages as one agent supplying information or

requesting help from another. Viewed this way, the main contribution of OO is two-fold: It highlights the value of structuring knowledge in relatively self-contained modules, and of organising it in abstract hierarchies.

Biconditionals. This chapter explores the view that conditional beliefs are two-way biconditionals in disguise. Both negation as failure and abduction can be understood as reasoning with such biconditionals as equivalences.

Computational Logic and the Selection Task. This chapter interprets some results of psychological experiments on reasoning with conditionals. It investigates the different ways CL explains these results, depending on whether a conditional is interpreted as a goal or as a belief. If it is interpreted as a belief, then it is often natural to interpret the conditional as specifying the *only* conditions under which the conclusion holds. This explains one of the two main mistakes that people make when reasoning with conditionals, when judged by the standards of classical logic. The other mistake is that people often fail to reason correctly with negation. This is partly explainable by the fact that an agent's observations are normally represented by positive atomic sentences, and that negative conclusions must be derived from positive observations. Often this is easier with conditional goals than with conditional beliefs.

Meta-logic. Meta-logic can be used to simulate the reasoning of other agents, and to solve problems that cannot be solved in the object language alone. This is illustrated with a variant of the wise man puzzle and, moreover, with Gödel's theorem that there are true but unprovable sentences in arithmetic.

Conclusions. This concluding chapter takes a step back from the details, and takes a broader look at the main aim of the book, which is to show how CL can reconcile conflicting paradigms for explaining and guiding human behaviour. It also suggests how CL may help to reconcile conflicts in other areas.

Appendices: The Syntax of Logical Form. Truth. Forward and Backward Reasoning. Minimal Models and Negation. The Resolution Rule of Inference. The Logic of Abductive Logic Programming.

Commentary

Robert Kowalski is a foundational figure of Computational Logic and, over the years, has regularly laid down in print a succession of its cornerstones. His latest is a book several years in the making, that has filled the community with justified expectation and reward.

It is a deep epistemological book, as can be amply gleaned from the detailed summary above, and one which unrelentingly makes the case for Computational Logic as a premium scaffolding framework that bridges computational and human thinking.

The book is within clear grasp of a general higher-educated audience, because of the adept and informal naturalness with which it addresses, explains and exemplifies nevertheless non-trivial issues in knowledge representation and reasoning.

It is also a treasure trove for teachers and researchers alike, as it admirably integrates the author's longstanding groundbreaking and fertile research efforts, and expounds with clarity and simplicity the unifying epistemological virtues of the Computational Logic paradigm – one that is supported by a vast community of researchers.

Additionally, the formal support material, concentrated in the 64-pages of the six appendices at the end, provides a self-contained introduction to Computational Logic, which furthermore pinpoints and raises problems with a high potential for subsequent research, of interest to the community as a whole.

The Conclusions, when all is said and done, stand convincingly vindicated, and the book's set of beliefs, achievement goals and maintenance goals are perfunctorily discharged, notwithstanding the author's own penitence on topics left out, which we countenance: namely Inductive Learning, Uncertainty, Probability Theory, Connectionism, Conflict Resolution, and their relationships to Computational Logic.

This is understandable given the general audience targeted, and the technicalities of such advanced subjects, it being justly compensated for by due reference to appropriate surveys and the scientific literature.

One could readily think of other similarly refractory advanced subjects, such as Ontologies, Preferences, Belief Revision, to name a few. Indeed, given the very richness and varied scope of the Computational Logic paradigm, it will remain a never ending source and destination for the modelling of human thinking and its enabling mechanisms.

The book's subtitle, "**How to be Artificially Intelligence**", appears at first puzzling and I believe deliberately so. Is natural intelligence not enough? Is that not for computers and the like? How can I, a human, be artificially intelligent?

Answers are several in coming, but all under the banner that to understand – and thence to perfect – human intelligence, one must express how it works, and the computer is the experimental apparatus to test our own self-understanding, modelling it in detail to the extent we can. Logic, in the wide sense of logical, is the natural shared vehicle for so doing in a precise scientific way. And the computer, our privileged computational machine *par excellence*, is no doubt our artificial shared vehicle for objectively proving the worthiness of that understanding. Computational Logic spans both, and symbiotically benefits both.

Robert Kowalski is indeed intelligent, whether artificial or not, so it is a great thing that he wrote this illustrious book, as only he could have done.