Midway through this account of the formal representation of reasoning agents acting in a changing environment, Michael Wooldridge writes that "all the formalizations of mental states that appear in this book… represent idealizations of true states of affairs" (115). It is no doubt true that *Reasoning about Rational Agents* contains idealizations. On the one hand, Wooldridge is a logician, and the complex-but still simplified-formal framework he uses to describe "being rational" necessarily elides some of the fine details of agents in the real world. On the other hand, he is a computer scientist, concerned with artificial agents that are intuitively based upon but nonetheless abstracted from the human case. Since the intended audience for the book will be familiar with, or at least interested in, formal approaches to artificial agency, such idealizations are not troubling. The real question, however, concerns the "true states of affairs" that Wooldridge's formulations are supposed to capture. It is here that a more general audience-philosophers and psychologists, for instance-will have more about which to argue.

Wooldridge's topic is LORA, the "Logic of Rational Agency," and readers will need to be comfortable with formal logic. That said, the presentation is not overly difficult, and anyone familiar with basic modal logic or the logics of time and agency will have little trouble following along. The book presents the state of the art in one approach to dealing with questions of artificial agency and reasoning, namely the "Belief-Desire-Intention" (BDI) model. This approach is "representationalist," taking BDI's to be sets of discrete items "possessed" by agents, with their contents describable in explicit terms, and with such as "deliberation" or "means-end reasoning" describable by way of functions from one such set to another. This approach is not shared by all those doing research in artificial agency, and those favoring an "inexplicit" or connectionist paradigm will likely not agree with much of what gets done here. Too, those concerned with actually implementing agency will not be entirely satisfied-as Wooldridge indicates, the semantics for LORA make machine implementations difficult. He does provide some hints about future directions this project could take, but points out that LORA is presently intended as a "specification" of rational agency, describing how rational agents ought to behave, though unable to show how to construct them. Indeed, one of the book's admirable features is that Wooldridge never ignores where further work needs to be done, never claiming that his approach answers all questions concerning rationality.

That said, the argument could still use some development. As mentioned, there is plenty here to contest. As an instance, consider the analysis of belief. Treating belief as a logical modality similar to the familiar necessity operator, Wooldridge links rationality to a form of logical perfection. Rational agents, on this view, believe all the consequences of their beliefs, and these beliefs are always perfectly consistent. This is doubtless an idealization, since few real-world agents of any real complexity are likely to be completely consistent in their beliefs. Furthermore, requiring that an agent believe all consequences of her beliefs is puzzling, especially if beliefs are discrete and explicit items. On such a view, agents possess an *infinite number* of beliefs and this, along with the existence of "infinitary belief-states" in instances of mutual belief, is difficult to reconcile with Wooldridge's stated intention that his view is meant to reflect realistic presumptions about the limited capacities of finite agents. Readers will have many such occasions to ask whether this logical representation of rationality is indeed a good one. The one main criticism is that Wooldridge does very little to address such worries, too often presenting his logical framework without discussing its meaning or defending its adequacy in any real detail. Given the complexity of his subject-matter, and the variety of opinions on it, this can be somewhat frustrating.

Still, the formal approach has the advantage of making its actual consequences clear, and a reader can quickly see what Wooldridge's theory entails, whether or not she agrees with him. Furthermore, someone looking for an
introduction to a complex area of research would be well advised to start with this book. Wooldridge provides an extensive bibliography, copious notes, and each chapter ends with a section pointing toward further reading. The book is a fine guide to the current state of thinking on the topic, and it is to be presumed that many of the questions it leaves open can be addressed, if not decisively answered, by following up the leads it lays down.