

# Comments on “Knowledge, Representation, and Rational Self-Government”

Position Paper

Hector J. Levesque  
University of Toronto

I think I agree with most of what Doyle says in his paper [Do88]. He begins by distinguishing between explicit and implicit belief, a distinction I also pursued in a 1984 paper [Le84]. He characterizes the former as “entries in an agent’s memory,” and the latter as the logical entailments of the former. He then goes on to argue how this notion of implicit belief is an inappropriate characterization (even from an idealized, competence viewpoint) of rational inference. I fully agree: the way to talk about inference is not in terms of a logic of implicit belief, but in terms of principles of rationality, problem-solving, and meta-reasoning, and with specific reference to the goals, intentions, preferences, and resource constraints of the reasoner in question.

So where do we disagree? First of all, I think he dismisses the notion of explicit belief too quickly. Certainly what I described in my paper was not merely a collection of memory items. The notion I had in mind was more like the beliefs that are architecturally available to the agent without further introspective problem-solving or meta-reasoning. The assumption is that no matter how reflective an agent can be, ultimately, there must be beliefs that can be obtained without further introspection, or reasoned control. The issue then is whether or not this base level of beliefs makes any semantic sense. In Doyle’s case, there would certainly be no logic of explicit belief, any more than there could be a logic of data structures. But in my case, I wanted to interpret explicit beliefs in terms of a (perhaps inconsistent, likely incomplete or disjointed) picture of the world, requiring further elaboration and refinement. So the set of explicit beliefs had to be semantically coherent, even if not closed under logical consequence. The logic I proposed was simply a characterization of (one version of) this coherence and absolutely not, as Doyle suggests, a non-classical logic for implicit belief. The reason I feel that a logical (rather than an “economic”) characterization of these beliefs is suitable is that by my definition, their management is not a problem that the agent has to think about.

What about implicit beliefs? As I said above, I don’t think the set of logical

consequences of the explicit beliefs appropriately describes the result of even idealized reasoning. But even if it does not contribute directly to the actions of a reasoning agent, I do think a logical characterization of these implicit beliefs is useful. Perhaps the simplest application is that agents need beliefs about these implicit beliefs to decide what to think about. In some cases, an agent may believe that the answer to a question simply requires more thought (for example, solving a mathematical puzzle); in other cases, an agent may believe that the answer is not even implicit in what she knows, and that further information is necessary. In other words, agents will normally need to have some picture (however incomplete) of what their implicit beliefs are like; a logic of implicit belief then, is simply a precise characterization of what they really are like.

[Do88] Doyle, J.: Knowledge, Representation, and Rational Self-Government. *This volume.*

[Le84] Levesque, H.: A logic of implicit and explicit belief. *Proc. Nat'l Conf. on Artificial Intelligence (AAAI-84)*, Austin, Texas, 1984, pp. 198–202.