Artificial Intelligence 223 (2015) 1–26

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Semantical considerations on multiagent only knowing

Vaishak Belle [a,*], Gerhard Lakemeyer [b]

[a] Dept. of Computer Science, University of Toronto, Toronto, Ontario, M5S 3H5, Canada
[b] Dept. of Computer Science, RWTH Aachen, 52056 Aachen, Germany

A R T I C L E I N F O

Article history:
Received 22 March 2013
Received in revised form 28 November 2014
Accepted 6 February 2015
Available online 12 February 2015

Keywords:
Knowledge representation
Multiagent systems

A B S T R A C T

Levesque introduced the notion of "only knowing" to precisely capture the beliefs of a knowledge base. He also showed how only knowing can be used to formalize nonmonotonic behavior within a monotonic logic. Levesque's logic only deals with a single agent, and therefore, a number of attempts have been made to generalize only knowing to the many agent case. However, all these attempts have some undesirable features. Most significantly, these attempts are propositional and it is not clear how they are to be extended to the first-order case. In this work, we propose a new semantical account of multiagent only knowing which, for the first time, has a natural possible-world semantics for a quantified language with equality. Among other things, properties about Levesque's logic generalize faithfully to the many agent case with this account. For the propositional fragment, we also provide a sound and complete axiomatization. Finally, we obtain a multiagent first-order version of the nonmonotonicity exhibited by the logic of only knowing.

1. Introduction

When considering knowledge-based agents, it seems natural to think that the beliefs of the agent are precisely those that follow from the assumption that the knowledge base (KB) is *all* that is known.[1] Think, for example, of an agent whose KB has a secret p, and nothing else. The agent, then, should not only infer that she knows a secret, which is p, but also that she does not know of any other secrets. In some situations it may be possible to fully specify the beliefs and non-beliefs of an agent, in which case classical epistemic logic suffices [1]. But in general, as has been argued elsewhere [2,3], a perspicuous characterization of the agent's epistemic state is very useful. This is especially significant from the perspective of a designer of some putative agent who would be representing the agent's beliefs in terms of a KB.

Levesque [4] was among the first to capture such a notion explicitly in his logic of "only knowing." His proposal, the logic $\mathcal{OL}$, is very simple. He augments a logic of belief [5,6], where (say) the modality $\boldsymbol{K}$ denotes knowledge, with a modality $\boldsymbol{O}$ to capture the notion of only knowing. Beliefs are reasoned about in terms of valid sentences of the form:

$$\boldsymbol{O}\,\mathrm{KB} \supset \boldsymbol{K}\alpha$$

* Corresponding author.
E-mail addresses: vaishak@cs.toronto.edu (V. Belle), gerhard@cs.rwth-aachen.de (G. Lakemeyer).

[1] For this work, the terms "knowledge" and "belief" are used interchangeably, with the understanding that knowledge need not necessarily be true in the real world.

http://dx.doi.org/10.1016/j.artint.2015.02.005

which is to be read as "if KB is all that is believed by the agent, then the agent knows α." What is particularly interesting about the new modality is that it not only allows one to draw conclusions about what is known but also about what is not. For example, $\boldsymbol{O}p \supset \neg\boldsymbol{K}q$ and, by introspection, $\boldsymbol{O}p \supset \boldsymbol{K}\neg\boldsymbol{K}q$ both come out valid. Note that this is quite different from classical epistemic logic, in the sense that if we replace $\boldsymbol{O}$ by $\boldsymbol{K}$, then neither of these sentences is valid. Thus, an agent is able to reason about its own ignorance, without having to be told explicitly what it does not know. Remarkably, the new modality has very intuitive (and simple) truth conditions: roughly a semantics is specified wrt a set of worlds e, where worlds are truth valuations; a KB is only known just in case it is true in all worlds in e and one cannot add more worlds without falsifying at least one of the formulas of the KB. These ideas have since been applied to a number of knowledge representation problems, including, for example, reasoning about dynamical systems [7].

When the KB itself refers to an agent's beliefs, then only knowing also exhibits a form of nonmonotonic reasoning. For example, consider the infamous Tweety default formulated as the following sentence δ:

$$\forall x.\ (Bird(x) \wedge \neg\boldsymbol{K}\neg Fly(x)) \supset Fly(x)$$

which says that birds fly unless believed otherwise. Then, $\boldsymbol{O}(Bird(Tweety) \wedge \delta) \supset \boldsymbol{K}Fly(Tweety)$ is a valid sentence in Levesque's logic, and so Tweety is formally off the ground (as desired).

Readers may note that the nonmonotonicity exhibited by only knowing feels similar in spirit to Moore's autoepistemic logic (AEL) [8]. Levesque substantiated this by proving that Moore's stable expansions of a KB α, which are defined meta-theoretically using fixpoints, correspond precisely to the beliefs of an agent who only knows α. Thus, a reconstruction of a major branch of nonmonotonicity happens entirely within a *monotonic* logic. Levesque also proposed an axiomatization that characterizes the semantics in the propositional case, which allows, among other things, formal proofs of its flavor of nonmonotonicity. In his axiom system, to better understand the relationship between $\boldsymbol{O}$ and $\boldsymbol{K}$, he interprets $\boldsymbol{O}\alpha$ to syntactically denote $\boldsymbol{K}\alpha \wedge \boldsymbol{N}\neg\alpha$ using a new modal operator $\boldsymbol{N}$. Here, $\boldsymbol{N}\neg\alpha$ is to be read as "at most α is known," which can be contrasted with the usual $\boldsymbol{K}\alpha$ that, in fact, could be read as "at least α is known" since $\boldsymbol{K}\alpha$ does not preclude $\boldsymbol{K}(\alpha \wedge \beta)$ from holding, in general. With this new operator in hand, Levesque shows that by treating $\boldsymbol{K}$ and $\boldsymbol{N}$ to be fully introspective, the following axiom is all that is needed to completely characterize the semantics:

A5. $\boldsymbol{N}\alpha \supset \neg\boldsymbol{K}\alpha$ if $\neg\alpha$ is a consistent propositional formula.

This is clearly a novel axiom, and while it gives only knowing the intended properties, it also raises concerns when considering generalizations (as we shall shortly see). We remark that Levesque also proposed an axiomatic characterization for the first-order case, but this was shown to be incomplete by Halpern and Lakemeyer [9]. In fact, Halpern and Lakemeyer proved a stronger result: there cannot be a complete and recursive axiomatization.

Levesque's logic, however, only deals with a single agent. It seems perfectly natural to consider only knowing in the context of multiple agents. For example, imagine Alice having a secret p. It is reasonable to imagine that she believes the following:

Unless I know that Bob believes a secret, he does not know it.

If this sentence and p is all that Alice knows, then it should come out that Alice believes that Bob does not know her secret p. More generally, of course, one would often like to perspicuously characterize the beliefs of agents in multiagent settings, and here too, an account of only knowing would prove useful.

Motivated by such scenarios, a number of attempts have been made over the years to extend only knowing to the many agent case [10–14]. That is, in contrast to classical epistemic logic where multiagent extensions are immediate [1], generalizing only knowing has been shown to be very challenging [12]. In a sense, this is because an interesting technicality: the modality $\boldsymbol{O}$ has a subtle relationship to the classical operator $\boldsymbol{K}$ which makes extensions non-trivial. Moreover, proposing an accompanying axiomatization is also challenging: notice that the multiagent version of **A5** cannot be restricted to propositional formulas, since this would preclude us from performing derivations with nestings of only knowing operators. Unfortunately, this leads to a circularity: we would like such an axiom to hold for any consistent (not necessarily objective) formula, but to deal with consistency we would need an axiom system.

We discuss the aforementioned attempts in greater detail in the subsequent sections, but basically, we show that all of them have undesirable properties:

- Early propositional proposals by Lakemeyer [10] and Halpern [11] are shown to exhibit unintuitive features about only knowing. In fact, in later work, Halpern and Lakemeyer [12] argue why neither of these proposals satisfactorily capture only knowing.
- To remedy some of the shortcomings of their proposals, Halpern and Lakemeyer [12] developed a revised account that does seem to successfully model multiagent only knowing. But it has proof-theoretic notions in the semantics. Precisely for this reason, that proposal is not natural. Most significantly, it is not clear how their ideas would generalize to the first-order case. That is, as we already noted above, there cannot be a complete first-order axiomatization of only-knowing even for a single agent. Hence we cannot even get off the ground envisaging a semantics that appeals to its own proof theory as in [12]. Finally, for generalizing **A5**, they define the semantic notion of validity directly in the language as a modal operator, which adds considerable complexity to the axiomatization for technical reasons only.