STUDIES IN LOGIC, GRAMMAR AND RHETORIC 30 (43) 2012

Dariusz Surowik
University of Bialystok

SOME PROBLEMS IN REPRESENTATION OF KNOWLEDGE IN FORMAL LANGUAGES

Abstract. In the article we discuss the basic difficulties which we have when we use a formal language to describe knowledge of an agent. In particular we discuss the difficulties if we accept positive and negative introspection of an agent, Fitch's paradox, the problem of logical omniscience and some ways to avoid them.

Keywords: knowledge, epistemic logic, omniscience, Fitch's paradox, introspection.

Some of the axioms of $S5$ epistemic logic are controversial, or lead to results that are regarded in some cases as undesirable. We will now describe the controversy and some problems that have been considered in the case of some axioms of modal epistemic logic.

Negative introspection of an agent

Real agents have limited cognitive self-awareness of knowledge. Depending on the type of agent, the agent may know that he knows something or does not know it. By accepting axioms which express the introspection of an agent we also accept the self-reflexivity of knowledge of this agent. However, the self-reflexivity of knowledge of the agent is at the core of various paradoxes.

Let us consider, for example, the difficulties which we have if we describe the knowledge of an agent who has knowledge about only one atomic sentence p. We assume that the agent is the ideal cognitive agent and he knows all the logical consequences of his knowledge (in the discussed case, the agent knows all the logical consequences of the sentence p) and he can be introspective of his knowledge. The question is whether knowledge about p and the logical consequences from p constitues the whole knowledge of our

ISBN 978–83–7431–362–9 ISSN 0860-150X

agent. Intuitively, it seems that the proper answer is *yes*. Let us assume that q is an atomic sentence different from p. The agent, in accordance with the accepted premise in the introduction, has knowledge about p only and he has no knowledge about q. So, we can conclude that $\neg Kq$. We have assumed, however, that the agent is the ideal cognitive agent and can be introspective of his knowledge. Thus, if the agent has knowledge of his ignorance about the sentence q, then $K\neg qK$. Although we have assumed that the agent has knowledge only about p, he also knows that $\neg Kq$. But $\neg Kq$ is not a logical consequence of p.

We have a more interesting case if we consider more than one agent. To distinguish "old" agent and "new" agent, let us assign the indices 1 and 2 respectively. Since by our assumption, agent 1 has only knowledge about p, so it is true that $\neg K_1 q$. Agents have knowledge only about true sentences. Therefore, since the sentence $K_1 q$ is false, agent 2 does not know that $K_1 q$, then $\neg K_2 K_1 q$. Agent 1 knows that agent 2 does not know that agent 1 knows that q, so $K_1 \neg K_2 K_1 p$. Agent 2 can also be introspective of his knowledge. So it is true that agent 1 knows that agent 2 knows that agent 2 does not know that agent 1 knows that q. Formally, we can write this as follows: $K_1 K_2 \neg K_2 K_1 p$. Agent 1, although his knowledge was limited to p, knows a nontrivial fact about the knowledge of agent 2.

The difficulties that have arisen in the considered example, are due to the negative introspection of agents. If we reject the negative introspection of agents, the model which we have used to characterize the knowledge becomes simpler and free of some contradictions.

Fitch's paradox

Fitch in [4] argued that if there are unknown truths, then there are also unknowable truths. The argument presented by Fitch sparked a lively and long discussion among philosophers and epistemologists. We will present a reconstruction of the reasoning described by Fitch.

Allow that p is true and unknown. We can write this as $p \wedge \neg Kp$. If $K(p \wedge \neg Kp)$, then if we use for it a rule $K(\varphi \wedge \psi) \vdash (K\varphi \wedge K\psi)$, we get $Kp \wedge K\neg Kp$. So, the sentences Kp and $K\neg Kp$ are true. However, using for the second sentence a rule $K\varphi \vdash \varphi$ we get $\neg Kp$. We have shown that under the assumption that $K(p \wedge \neg Kp)$ the sentences Kp and $\neg Kp$ are inferable. It can not be that $K(p \wedge \neg Kp)$, because this leads to a contradiction.

Let us assume that $\neg K(p \wedge \neg Kp)$. If we apply Gödels's rule to this formula we obtain $\Box\neg K(p\wedge\neg Kp)$. The operators of necessity and possibility are definable for each other ($\Box p \equiv \neg\Diamond\neg P$) so we get $\neg\Diamond\neg\neg K(p \wedge \neg Kp)$. Hence, from the double negation law and rules of substitution we get:

$$\neg\Diamond K(p \wedge \neg Kp).$$

The above formula says that it is impossible to know that $p \wedge \neg Kp$. So, if p is true and unknown, then the fact that p is true and unknown is unknowable.

If we accept the principle that *if something is true, then it is possible to know it* ($\varphi \rightarrow \Diamond K\varphi$), then if we assume that $p \wedge \neg Kp$, from the above principle and Modus Ponens we get:

$$\Diamond(p \wedge \neg Kp).$$

This is in contradiction with the previously obtained formula. Since the assumption that $p \wedge \neg Kp$ leads to a contradiction (regardless of whether we believe that this formula is known or unknown), so we assume that $\neg(p \wedge \neg Kp)$. From the last formula and the thesis of classical logic, we have that $p \rightarrow Kp$, which can be interpreted as: if p is true, then p is known. The assumption that $p \wedge \neg Kp$ and the commonly accepted rules leads to a contradiction. The rejection of this assumption leads to the conclusion that if something is true, it is known. We made no assumption on p, so from the above considerations we conclude that all truths are known. This is, of course unintuitive and inconsistent with reality. We do not know all truths and we have not been bestowed with the property of omniscience.

The above reconstruction of Fitch's argument could not be carried out if we had not established the acceptability of two rules, essential for our reconstruction: the rule of the distributivity of the knowledge operator with respect to the conjunction operator $K(\varphi \wedge \psi) \vdash (K\varphi \wedge K\psi)$ and the rule of the factuality of knowledge $K\varphi \vdash \varphi$. It seems that the rejection of these two rules is sufficient to ensure that Fitch's argument can not be reconstructed. Such an approach, however, is an unsatisfactory approach for a few reasons. Firstly, both rules have good reasons in knowledge systems. The rule of factuality of knowledge has been widely discussed as a prerequisite of knowledge. The rule of distributivity is quite intuitive and if we reject this rule we obtain a notion of knowledge with very specific properties. Secondly, the rules are not necessary for the reconstruction of Fitch's argument. Tennant in [12] and [13] shows that it is possible to reconstruct Fitch's argument without involving the rule of factuality of knowledge. Williamson in [15] formulate a version of the principle of intelligibility such that it is possible to derive inconsistent consequences without refering to the distribution of knowledge operator with respect to the conjunction operator[1].

Omniscience problem

The best known way to formalize knowledge and belief is an epistemic modal logic. Modal languages are a combination of the high power of expression and intuitive semantics – the reason they are a great tool used in artificial intelligence. The main achievement of modal logics is the transformation of extensional languages to intensional languages.

Unfortunately epistemic modal logic suffers from an ailment called *logical omniscience*. The problem of logical omniscience makes formal systems which use epistemic logic not a proper tool for modeling real agents, because real agents are never logical omniscient. The basic axiom of normal modal logics $K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$ leads to the conclusion that agents knows all logical consequences of their knowledge. This assumption is unreasonable in relation to the real agents (eg. people) or agents with bounded memory (eg. processors in computer systems). If we assume, for example, that an agent knows the algorithm of how to decompose a number into its prime factors, then the agent with common assumptions being made in the formal system, used to describe the knowledge, should to know the decomposition into its prime factors of all natural numbers. It's hard to defend that assumption with respect to the real agents because we do not expect that a real person has a knowledge of the decomposition into prime factors of all natural numbers, which might have, for example, five hundred thousand digits. Modal knowledge operators are regarded as an idealization of knowledge operators used in reasoning by humans.

In order to reject the disputed ownership of agents, such as their logical omniscience, there are some different modifications of systems of modal logic.

Konolige's suggestion

One of the possible methods used to reject the logical omniscience of agents is the syntactic approach, in which is assumed that the knowledge of agents is represented by an arbitrary set of formulas [9], [3]. Of course, the set

[1] Fitch's argument was also discussed on the basis of intuitionistic logic [14] and [1].

we are talking about should be constructed in such a way that it is closed on the logical consequence and should include the set of all substitutions of all axioms. The notion of knowledge formulated above does not lead to the problem of logical omniscience paradox, however, knowledge understood in this way is very difficult to analyze. If knowledge is represented by an arbitrary set of formulas, we do not have any guidance on how to analyze such knowledge. Konolige [6] gives a proposal for constructing such a set. In Konolige's approach, knowledge of an agent is represented by a set of primitive facts, which is closed to the rules of inference. In this case, the problem of logical omniscience is solved by the adoption of an incomplete set of inference rules.

Montague's semantics

Montague in [8] gives possible worlds semantics for epistemic logic, in which formulas are associated with sets of possible worlds, but knowledge is not modeled as a relation between possible worlds [16]. There is no relation of accessibility in Montague's semantics. In this approach we lose the intuition that the agent knows that φ, if φ is true in all worlds considered by the agent as possible. However, in this approach we can avoid the problem of logical omniscience. There is some difficulty in Montague's semantics. Namely, in this semantics, agents do not know all the logical consequences of their knowledge, so they are not able to identify formulas which are logically equivalent.

Impossible worlds

There have been efforts of trying to avoid the problem of logical omniscience by enriching the standard semantics of possible worlds with impossible worlds. An impossible world is a world in which the proved formula may not necessarily be true, or a world in which true formulas are contradictory. Impossible worlds are worlds that only enrich the epistemic set of epistemic alternatives, however, they are not logically possible worlds. With this approach, agents may not know all the tautologies of classical logic, since there may be possible worlds, considered by the agents as possible, in which some tautologies are not true.

The proposal of a semantics which contains impossible worlds was given by Levesque in [7]. Levesque makes a distinction between *explicit knowledge*

(knowledge provided *outside*) and *implicit knowledge* (knowledge that consists of all the logical consequences that can be derived from explicit knowledge). Levesque considered a model of possible worlds in which atomic sentences can be true, false, true and false at the same time, and such that they have no specified property. Levesque gives a logic of *internal* (implicit) and *external* (explicit) beliefs. Levesque's results can be extended to the case of knowledge [5, p. 8]. Explicit beliefs imply an implicit belief, but not vice versa. In Levesque's logic there is no problem of logical omniscience because in this logic there is no rule: if φ is a tautology, then there is $B\varphi$[2].

Semantics, which allow the existence of impossible worlds are also described for example in [2], [10], [11].

REFERENCES

[1] Beall J. C. *Fitch's proof, verificationism, and the knower paradox.* Australasian Journal of Philosophy, 78:241-247, 2000.

[2] Cresswell M. J., *Logics and Languages*, Methuen and Co., 1973.

[3] Eberle R. A., *A logic of believing, knowing and inferring*, Synthese 26, 1974, pp. 356–382.

[4] Fitch F., *A Logical Analysis of Some Value Concepts*, The Journal of Symbolic Logic, 28, 1963, pp. 135–142.

[5] Halpern J. Y., *Reasoning about knowledge. An overview*

[6] Konolige K., *Belief and incompleteness*, SRI Artificial Intelligence Note 319, SRI International, Menlo Park, 1984.

[7] Levesque H., *A logic of implicit and explicit belief*, Proceedings of the National Conference on Artificial Intelligence, 1984, pp. 198–202; a revised and expanded version appears as FLAIR Technical Report 32, 1984.

[8] Montague R., *Universal grammar*, Theoria 36, 1970, pp. 373–398.

[9] Moore R. C. and Hendrix G., *Computational models of beliefs and the semantics of belief sentences*, Technical Note 187, SRI International, Menlo Park, 1979.

[10] Rantala V., *Impossible worlds semantics and logical omniscience*, Acta Philosophica Fennica 35 1982, pp. 106–115.

[2] Levesque showed that for a certain class of his logic formulas, namely formulas of the form $B\varphi \to B\psi$, where φ and ψ are sentences in conjunctive normal form, the problem of finding the truth about the character formula is decidable in polynomial time. Thus makes Levesque's logic can be seen as a useful tool for modeling the concept of knowledge proposed by Levesque.

[11] Rescher N. and Brandom R., *The Logic of Inconsistency*, Rowman and Littlefield, 1979.

[12] Tennant N. *The Taming of the True.* Oxford Univeristy Press, USA, 1997.

[13] Tennant N. *Is every truth knowable?* reply to Hand and Kvanvig. Australasian Journal of Philosophy, 79:107–113, 2001.

[14] Williamson T. *Intuitionism disproved?*, Analysis, 42:203–207, 1982.

[15] Williamson T. *Knowledge and Its Limits.* Oxford Univeristy Press, USA, 2000.

[16] Vardi M. Y. 1986, *On epistemic logic and logical omniscience*, Proceedings of the Conference on Theoretical Aspects of Reasoning About Knowledge (ed. J. Y. Halpern), Morgan Kaufmann.

Dariusz Surowik
Chair of Logic, Informatics and Philosophy of Science
University of Bialystok, Poland
surowik@uwb.edu.pl