WHAT EPISTEMIC MODELS CANNOT BE

YANG LIU

1.1. The basic idea presented in Hintikka's 1962 book *Knowledge and Belief*, which has since become the standard approach in epistemic logic, was to treat knowledge as a modal operator with a semantic interpretation that an agent is said to know certain proposition if the proposition is true in all the worlds she considers *epistemically possible*. This approach, it is said, does not purport to provide a substantive account of knowledge, but to demonstrate a logic of the propositional attitude 'knows that.' The principal assumption in Hintikka's semantic theory is that our attributions to the propositional attitude in question are bounded by our abilities to exclude from all possibilities the ones that are considered epistemically inaccessible; hence what we know in any given situation is completely determined by those alternative epistemic possibilities that are compatible with our relevant attitudes (Hintikka, 1969). In this short paper, we remark on a puzzling phenomenon in this this way of modeling knowledge. More precisely, we show that within the standard possible world semantics the following cannot simultaneously hold:

i. The agent is aware of her current epistemic state.
ii. The agent has access to her own epistemic structure.
iii. The agent is an **S4** agent.

1.2. To illustrate, some technical preliminaries are needed. Let Ω be a non-empty set of possible worlds and $\rightsquigarrow$ be the *epistemic alternativeness* relation of the agent in question, which is a binary relation defined on $\Omega \times \Omega$: for any $\omega, \nu \in \Omega$, '$\omega \rightsquigarrow \nu$' says that ν is considered by the agent in world ω as an epistemic alternative. Let us denote by $\mathcal{K}(\omega)$ the set of all epistemic alternative worlds at ω, namely,

$$\mathcal{K}(\omega) = \{\omega' \in \Omega \mid \omega \rightsquigarrow \omega'\}. \tag{1}$$

We also refer to $\mathcal{K}(\omega)$ as the **epistemic state** of the agent at ω. By the agent's **epistemic structure** (or *knowledge structure*) $\mathcal{K}$ we mean the set of all epistemic states, that is,

$$\mathcal{K} = \{\mathcal{K}(\omega) \mid \omega \in \Omega\}. \tag{2}$$

It is plain that the agent's epistemic structure provides a description of her epistemic states in all possible worlds. By definition, a proposition p is said to be known by the agent at ω, denoted by $\omega \vDash Kp$, if p is true in all the worlds in $\mathcal{K}(\omega)$. It is further assumed, as argued in length by Hintikka (1962), that the alternative relation $\rightsquigarrow$ is reflexive and transitive, which correspond, respectively, to the truth axiom ($Kp \to p$) and the positive introspection axiom ($Kp \to KKp$) in the underlying formal system, namely an **S4** logic of knowledge.

Date: September 23, 2013.

FIGURE 1. a model of **S4**

Presumably, in any given world ω, the agent is aware of her current epistemic state $\mathcal{K}(\omega)$ and, as we shall argue in greater detail in §1.5, it is further assumed that the epistemic structure $\mathcal{K}$ itself is knowable *a priori* to the agent in the current semantic theory. However we show by means of a simple example that these assumptions cannot be coherently maintained in an **S4** system.

1.3. Consider the example illustrated in Figure 1, where Ω contains two possible worlds and the agent's alternativeness relation is represented by the directed graph with $\mathcal{K}(\omega_1) = \{\omega_1, \omega_2\}$ and $\mathcal{K}(\omega_2) = \{\omega_2\}$. This obviously satisfies **S4**. Suppose that ω_1 is the real world and that p is true in ω_1 but false in ω_2. Then it is easily seen, by the truth definition, that p is not known by the agent at ω_1.

However, observe that, at ω_1, the agent's epistemic state is $\mathcal{K}(\omega_1) = \{\omega_1, \omega_2\}$, she might as well reason: "Given my epistemic structure $\mathcal{K} = \{\mathcal{K}(\omega_1), \mathcal{K}(\omega_2)\}$ where $\mathcal{K}(\omega_1) = \{\omega_1, \omega_2\}$ and $\mathcal{K}(\omega_2) = \{\omega_2\}$, the only possible world in which I could be in my current epistemic state $\{\omega_1, \omega_2\}$ is world ω_1, therefore ω_1 must be the real world." This consideration leads the agent to exclude ω_2 as an alternative. Hence, at ω_1, p becomes known. But this is all very puzzling because the change of the agent's knowledge of p from unknown to known rests on nothing but the mere reflection on her own epistemic structure!

The puzzle can be also stated in the following form: let Ann be any **S4**, but not **S5** agent and let p be any proposition, which Ann does not know, but also does not know that she does not know (i.e., $\neg K_a \neg K_a p$). For example, let p be expressed by the sentence "Harvey Lee Oswald had planned and executed Kennedy's assassination all by himself." Now suppose that Bill, who knows that Ann is an **S4** agent, asks Ann whether she knows that p. Assume moreover that the exchange is subject to the rule that the speaker will not assert any proposition that she does not know to be true. Furthermore, the agents are fully cooperative and will provide the fullest information they have that is relevant to the question. In that case, Ann cannot give any yes/no answer to Bill's question. Bill can then deduce from the fact that Ann could not answer 'Yes' that Ann does not know that p. But if Bill can deduce it, why cannot Ann herself deduce it? In fact, why can't Ann ask this question herself, "Do I know that p?" and then, if she knows that she is an **S4** agent, she can derive, from her inability to answer 'Yes,' the fact that she does not know. Thus she becomes an **S5** agent. If she knows that she is an **S4** agent, this deduction requires only a minimal meta-reflection on herself. The conclusion seems to be that an **S4**, but not **S5** agent, cannot know that she is an **S4** agent.[1]

1.4. Observe that the agent's reasoning above is in effect equivalent to defining the following *epistemic indistinguishability* relation among possible worlds: world ω is said to be indistinguishable from world ν for the agent, in symbols $\omega \sim \nu$, if and only if the agent is in exactly the same epistemic state at either of the two

[1]I thank Prof. Haim Gaifman for providing this example of Ann and Bob.

worlds, i.e., if and only if $\mathcal{K}(\omega) = \mathcal{K}(\nu)$. It is plain that $\sim$ is an equivalence relation under which the space Ω is divided into equivalence classes. In the example above, we have that $[\omega_1]_\sim = \{\omega_1\}$; in other words, the only world that is indistinguishable from world ω_1 for the agent is ω_1 itself, in which p is true and hence known by the agent.

But now we have two candidates for the epistemic *accessibility relation*, namely the alternativeness relation $\rightsquigarrow$ and the indistinguishability relation $\sim$ just defined, both of which can be said to representations of the agent's epistemic capacity to locate actuality throughout the logical space of possibilities. However, as shown in the example above. The two representations may exemplify entirely different structures. Then the question is which one should be the "correct" accessibility relation to be employed in an epidemic model?

One attempt to reconcile the two accessibility relations is to invoke the full **S5** logic, and it can be easily shown that the alternativeness relation and the indistinguishability relation coincide with one another in this logic. However this amounts to adopting, in the logic of knowledge, the negative introspection axiom ($\neg Kp \to K\neg Kp$), which has been explicitly rejected by Hintikka (1962, §3.8) as an unreasonable principle for knowledge.

1.5. To be sure, the crux of the puzzlement is this: Is the epistemic structure $\mathcal{K}$ itself knowable to the knower? For, as seen in the discussion above, the shift from the alternativeness relation $\rightsquigarrow$ to the indistinguishability relation $\sim$ relies on the assumption that the agent can reflect on her own epistemic structure, which allows the agent to distinguish those possible worlds that yield different epistemic states. It might be contested that this assumption is more than an innocuous presupposition on the part of the agent, for it essentially amounts to granting the agent an "outsider's view" (like the picture in Figure 1) over what she could in principle know or not know, which nonetheless is not characteristic of an **S4** agent (who does not know what she does not know). The skeptics may hence conclude that it is this first-person/third-person asymmetry that makes the assumption of epistemic structure impossible to maintain.

However, the semantic theory developed by Hintikka seems to suggest that the epistemic structure is knowable *a priori* to the knower. In an expository article Hintikka and Halonen put the following

> In what way are you better off when you come to know that S [a proposition]? The basic answer is clear. When you know that S, you can legitimately omit from consideration all possibilities under which it is not the case that S; in other words, you can restrict your attention to the situations in which it is true that S. (Hintikka and Halonen, 1998, §1)

In other words, it is a presupposition in their theory that the agents are aware of their own epistemic structure $\mathcal{K}$ and hence her epistemic state in any given world. For, otherwise, the agent, viewing from her first-person standpoint, will be completely confined to her current epistemic state, which prevents her from ever acknowledging the entire state space Ω, which may contain possible worlds that are not in the agent's current epistemic state $\mathcal{K}(\omega)$. But this only stands in contrary

to the assumption that the agent is capable of eliminating from *all* possibilities the ones that are considered epistemically impossible. Hence it seems that the assumption that the agent has knowledge of her own epistemic structure cannot be rejected without making fundamental changes in how knowledge is modeled in the present possible-world framework.

The first-person/third-person asymmetry mentioned above can also be viewed in the following way: there are different *kinds* of knowledge involved, namely, the knowledge *of* what the agent knows and the meta-knowledge *about* what the agent knows (i.e., the knowledge of agent's epistemic structure). In a semantic theory, the meta-knowledge about what the agent knows refers to the knowledge of the epistemic model itself. Then the question becomes: Can an epistemic model encodes the meta-theoretic properties of its own? Or, put differently, *Can a model of knowledge model the knowledge of the model*? If the meta-knowledge of what the agent knows is part of what a knower knows, or more so, as put by Hintikka, the meta-knowledge of the model be tautological in the underlying logic, then the first-person/third-person asymmetry vanishes. In this case, the agent may as well play the role of the theorist who sets out to model the epistemic capacity of her own. This however makes the modeling of an **S4** agent essentially impossible, for the acquisition of meta-knowledge of the model implies that the agent becomes aware of what she would not otherwise have known.

In sum, we are facing the following dilemma in Hintikka's logic of knowledge: If an agent is an **S4** agent but not **S5**, then it cannot be the case that the agent knows her own epistemic structure, in another word, it cannot be that she knows that she is an **S4** agent. On the other hand, if an agent knows her epistemic structure, then it cannot be the case that she is not an **S5** agent.

Acknowledgement. Thanks are due to Haim Gaifman, Isaac Levi, Rohit Parikh, Rush Stewart and Yanjing Wang for inspiring exchanges.

REFERENCES

Hintikka, J. 1962. *Knowledge and Belief: an introduction to the logic of the two notions*. Ithaca, New York: Cornell University Press.

Hintikka, J. 1969. Semantics for propositional attitudes. In *Models for Modalities: selected essays*, Synthese Library. D. Reidel Dordrecht.

Hintikka, J. and I. Halonen. 1998. Epistemic logic. In *Routledge Encyclopedia of Philosophy.*, ed. E. Craig. Routledge.

DEPARTMENT OF PHILOSOPHY, COLUMBIA UNIVERSITY, NEW YORK, NY 10027
E-mail address: yl2435@columbia.edu