From Pritchard & Bernecker 2011

52

LOGIC AND FORMAL SEMANTICS FOR EPISTEMOLOGY

John Symons

Epistemic expressions such as 'knows that' or 'believes that' have systematic properties that are amenable to formal study. Most obviously, statements containing epistemic expressions sometimes involve logical constants which behave in the usual way. So, for example, if you know *p* and *q* then you know *q*. The conceptual features of statements concerning knowledge and belief become more interesting when one begins to examine the characteristics of general principles governing the use of epistemic concepts. The behavior and interaction of these general principles has been the focus of epistemic logic. For example, as G.E. Moore pointed out, there seems to be something wrong with claiming

(1) "*p* and I do not believe *p*."

Assertions of this kind are self-defeating because of the conceptual features of knowledge or belief and not because of the syntactical features of the sentence or the character of the logical constants that are involved. As Moore noted "'I went to the pictures last Tuesday, but I don't believe that I did' is a perfectly absurd thing to say" (1952: 543). The *perfect absurdity* here is due to a violation of a principle governing epistemic concepts.

Notice that (1) is often a correct description of the state of affairs in question. For instance, since I recognize that I am fallible, I am committed to the possibility that there are cases where it is true that *p* and I do not believe *p*. Furthermore, I can assert, without paradox or contradiction, that there is some proposition *p* such that *p* and I do not believe *p*.

The paradox arises from the peculiarity of the agent in question attesting to particular instances of (1), where the variable *p* is replaced by an assertion concerning some state of affairs. Specifically, it is paradoxical insofar as it is, what John Austin called, an illocutionary act (1975: 133). While I recognize that there might be cases where replacing *p* with some description of some state of affairs is true, I cannot sincerely attest to both parts of the conjunction contained in (1) at a particular moment for any specific instance of (1). In this sense, Moore's paradox sheds light on the properties of epistemic agents and the concept of belief. Reflecting on the *perfect absurdity* of Moore's examples, shows us that an agent's belief and its agency are related. However, we are not restricted to relying on epistemic intuitions in our consideration of Moore's paradox. Exploration of the principles or norms governing epistemic notions can take place in an axiomatic fashion. Jaakko Hintikka, for example, provided a proof of the contradictory nature of the paradoxical form of the Moore statements, the case where the statement asserts that an agent believes p and *not-p* (1962: 67). I can take as a rule for instance that it is prohibited or confused to say:

(2) "I know p but it is not the case that p."

If (2) is prohibited, it is due in part to the illocutionary considerations which applied in (1) but unlike assertions of belief, (2) is prohibited by virtue of another general principle, namely the veracity of knowledge.

(3) If one knows p then it is true that p.

If we accept that knowledge implies veracity then we can consider what the implications of taking it as an axiom might be and whether it is consistent with other general epistemic principles we might hold. As discussed below, considerations of this kind have been given an elegant formal framework by epistemic logicians. Wolfgang Lenzen (1978) provided an excellent overview of arguments from the 1960s and 1970s concerning the appropriate axioms for knowledge.

Observations of the kind emphasized by Moore, concerning the behavior of the term "knows that" served as the starting points for the development of modern epistemic logic. G.H. von Wright was the first to sketch an axiomatic treatment of the behavior of epistemic concepts (1951: 29–35). However, modern epistemic logic began in earnest once Hintikka provided a semantic interpretation of epistemic and doxastic notions in the early 1960s.

Hintikka began by supplementing the language of propositional logic with two unary epistemic operators K_a and B_a such that $K_a p$ reads 'Agent a knows p' and $B_a p$ reads 'Agent a believes p' for some proposition p. In this way, candidate epistemic or doxastic axioms can be presented in formal terms. So, for instance, we have already seen that one intuitive axiom which we are likely to accept into our epistemic logic is:

(4) $K_c A \rightarrow A$

This is known as axiom T which we saw above as (3). With our modest addition to first-order logic in hand, we can begin to catalog other plausible epistemic axioms.

A standard list of the axioms (following Lemmon (1977), Bull and Segerberg (1984)) that are relevant for epistemic logic run as shown in Table 52.1:

Table 52.1 Axioms of Epistemic Logic

K	$K_c(A \rightarrow A') \rightarrow (K_c A \rightarrow K_c A')$
D	$K_c A \rightarrow \neg K_c \neg A$
T	$K_c A \rightarrow A$
4	$K_c A \rightarrow K_c K_c A$
5	$\neg K_c A \rightarrow K_c \neg K_c A$

.2 $\neg K_c \neg K_c A \rightarrow K_c \neg K_c \neg A$

.3 $K_c(K_c A \rightarrow K_c A') \vee K_c(K_c A' \rightarrow K_c A)$

.4 $A \rightarrow (\neg K_c \neg K_c A \rightarrow K_c A)$

We can consider the philosophical merits of each axiom to a certain extent without the introduction of additional formalism. However, Hintikka's approach to the semantics of epistemic notions offers an important supplement to our intuitive reflections. In order to begin thinking about the relative merits of these axioms one can begin by considering the familiar interpretation of the K and B operators using possible world semantics along the lines discussed above:

$K_c A$: *In all possible worlds compatible with what c knows, it is the case that A*
$B_c A$: *In all possible worlds compatible with what c believes, it is the case that A*

The basic assumption is that any ascription of propositional attitudes such as knowledge and belief, involves dividing the set of possible worlds in two: Those worlds compatible with the attitude in question and those that are incompatible with it.

The central idea in possible worlds semantics is the notion of accessibility. Accessibility is a relation that is defined on the set of possible worlds. In standard modal logic we say that some world w is accessible from some world w' just in case w is *possible* relative to w'. Specifically, the relation can be characterized as a subset of the Cartesian product of the set of possible worlds. As described below, determining the accessibility relation is the most basic step in determining the properties of our semantical framework. So, for example, whether one assumes that the accessibility relation is symmetric, transitive, reflexive, or some combination of the three, will make a significant difference in how one thinks about the modal or epistemic properties of the system in question. In the epistemic context, the set of worlds accessible to an agent (its set of epistemic alternatives) depends on its informational resources at an instant. This dependency is captured via the specification of the accessibility relation, R, on the set of possible worlds. To express the idea that for agent c, the world w' is compatible with his information state, or accessible from the possible world w which c is currently in, it is required that R holds between w and w'. This relation is written Rww' and reads "world w' is accessible from w". The world w' is said to be an *epistemic* or *doxastic alternative* to world w for agent c, depending on whether knowledge or belief is under consideration. We can give this a semantic interpretation, by saying that if a proposition A is true in all worlds which agent c considers possible then c knows A.

A possible world semantics for a propositional epistemic logic with a single agent c then consists of a *frame* $\mathcal{F}$ which in turn is a pair $\langle W, R_c \rangle$ such that W is a non-empty set of possible worlds and R_c is a binary accessibility relation (relative to agent c) over W. A *model* $\mathcal{M}$ for an epistemic system consists of a frame and a denotation function φ assigning sets of worlds to atomic propositional formulas. Propositions are taken to be sets of possible worlds; namely the set of possible worlds in which they are true. Let *atom* be the set of atomic propositional formulae, then $\varphi : atom \mapsto P(W)$, where P denotes the powerset operation.

The model $\mathcal{M} = \langle W, R_c, \varphi \rangle$ is called a Kripke-model and the resulting semantics Kripke-semantics (Kripke 1963): An atomic propositional formula, **a**, is said to be true in a world w in $\mathcal{M}$ (written $\mathcal{M}, w \models \mathbf{a}$) iff w is in the set of possible worlds assigned to **a**, i.e., $\mathcal{M}, w \models \mathbf{a}$ iff $w \in \varphi(\mathbf{a})$ for all $\mathbf{a} \in atom$. The formula $K_c A$ is true in a world w (i.e.,

$\mathcal{M},w \vDash K_c A$) iff $\forall w' \in W$, if $R_c ww'$, then $\mathcal{M},w' \vDash A$. The semantics for the Boolean connectives follow the usual recursive recipe. Similar semantics can be formulated for the belief operator. Since a belief is not necessarily true but, rather, probably true, possibly true, or likely to be true, we must modify our approach to the semantics of belief appropriately. For instance, belief can be modeled by assigning a sufficiently high degree of probability to the proposition in question and determining the doxastic alternatives accordingly. The truth-conditions for the doxastic operator are defined in a way similar to that of the knowledge operator and the model can also be expanded to accommodate the two operators simultaneously.

A modal formula is said to be *valid* in a frame if, and only if, the formula is true for all possible assignments in all worlds in the frame.

An important feature of possible world semantics is that the epistemic axioms listed above, correspond to algebraic properties of the frame in the following sense: A modal axiom is valid in a frame if, and only if, the accessibility relation satisfies some algebraic condition (see Hendricks and Symons 2006). For example, the axiom expressing the veridicality property that if a proposition is known by c, then A is true,

(5) $K_c A \rightarrow A$,

is valid in all frames in which the accessibility relation is reflexive in the sense that $\forall w \in W$: Rww. Given reflexive accessibility, every possible world is accessible from itself. Similarly if the accessibility relation satisfies the condition that

(6) $\forall w, w', w' \in W: Rww' \wedge Rw'w' \rightarrow Rww'$

which is also known as transitivity, then the axiom (7) is valid.

(7) $K_c A \rightarrow K_c K_c A$

(7) is called axiom 4 and is also known as the axiom of self-awareness, positive introspection, or the KK-thesis. In this case, the axiom captures the idea that if the agent knows p then it has knowledge of its knowledge that p. Other axioms require yet other relational properties to be met in order to be valid in all frames: If the accessibility relation is reflexive, symmetric and transitive, then

(8) $\neg K_c A \rightarrow K_c \neg K_c A$

is valid. (8) is called axiom 5, also better known as the axiom of wisdom. This is the much stronger thesis that an agent has knowledge of its own ignorance: If a does not know p, it knows that it doesn't know p. The axiom is also known as the axiom of negative introspection.

One contentious axiom which is valid in all possible frames,

(9) $K_c (A \rightarrow B) \rightarrow (K_c A \rightarrow K_c B)$,

is the closure condition for knowledge, also known as axiom K, or the axiom of deductive cogency: If the agent a knows $p \rightarrow q$, then if a knows p, a also knows q. As discussed below, this axiom leads to the most difficult philosophical problem for epistemic

logicians, namely the
accepts this axiom, th
from its knowledge.
Other axioms of e
in order to be valid ir
up epistemic modal
the modal formulas v
assumed for the acces
Returning to the a
their relative strengt
to be system T. The r
modal axiom K and
the axiom T from th
reflexive frame with
Additional modal
from the above pool,
way of example, whil

(10) $K_c A \rightarrow A$

is valid in system T,

(11) $K_c A \rightarrow A$,

are all valid in S5 bu
System T has a re
metrical accessibility
to which the arrow is
and hence reflect rel
ones listed.

Table 52.2 Relative

KT4
KT4 + .2
KT4 + .3
KT4 + .4
KT5

One of the importa
tems of such logics
ics range from S4 o
Hintikka settled fo
S4.2 (1978), van de
S4.3 (Meyer and va
together with Fagir
knowledge to be S5

logicians, namely the apparent commitment to logical omniscience. It seems that if one accepts this axiom, then an epistemic agent must know everything that follows logically from its knowledge.

Other axioms of epistemic import require yet other relational properties to be met in order to be valid in all frames. When combined in various ways, these axioms make up epistemic modal systems of varying strength. Their strengths vary according to the modal formulas valid in the respective systems and given the algebraic properties assumed for the accessibility relation.

Returning to the axioms listed above, we can begin to see how we might compare their relative strengths. The weakest system of epistemic interest is usually considered to be system **T**. The reader should take care to distinguish the epistemic operator *K*, the modal axiom K and the system of axioms **K** in what follows. Similarly, we distinguish the axiom T from the system **T**. T is a system of modal logic which is characterized by reflexive frame with the axioms T and K as valid axioms.

Additional modal strength can be obtained by extending **T** with other axioms drawn from the above pool, altering the frame semantics to validate the additional axioms. By way of example, while

(10) $K_c A \rightarrow A$

is valid in system **T**,

(11) $K_c A \rightarrow A$, $K_c A \rightarrow K_c K_c A$ and $\neg K_c A \rightarrow K_c \neg K_c A$

are all valid in **S5** but not in **T**.

System **T** has a reflexive accessibility relation, **S5** has reflexive, transitive and symmetrical accessibility relations. The arrows in Table 52.2 below indicate that the system to which the arrow is pointing is included in the system from which the arrow originates and hence reflect relative strength. Then **S5** is the strongest and **S4** the weakest of the ones listed.

Table 52.2 Relative Strength of Epistemic Systems Between S4 and S5

		Epistemic Systems	
KT4	=	S4	
KT4 + .2	=	S4.2	↑
KT4 + .3	=	S4.3	↑
KT4 + .4	=	S4.4	↑
KT5	=	S5	↑

One of the important tasks of epistemic logic is to catalog all sound and complete systems of such logics in order to allow us to pick the most 'appropriate' ones. The logics range from S4 over the intermediate systems S4.2–S4.4 to S5. By way of example, Hintikka settled for S4 (1962), Kutschera argued for S4.4 (1976), Lenzen suggested S4.2 (1978), van der Hoek has proposed to strengthen knowledge according to system S4.3 (Meyer and van der Hoek 1995). Van Ditmarsch, van der Hoek and Kooi (2007) together with Fagin, Halpern, Moses and Vardi (Fagin et al. 1995) and others assume knowledge to be S5 valid.

In the doxastic context, we can also catalog the completeness properties of the alternative systems in a similar fashion. Of course in doxastic logic we drop axiom T, which is usually replaced by D. This avoids committing doxastic logic to the truth of beliefs while retaining the condition that beliefs be consistent. Replacing T with D generates systems like KD4–KD45. This approach permits the combination of epistemic and doxastic systems and for studying the interplay between knowledge and belief (see Voorbraak 1993). There are some important philosophical concerns with such combined doxastic and epistemic systems. Lenzen (1978) and Stalnaker (1996) point out that such combined systems risk conflating knowledge and belief.

How does semantic formalization relate to epistemology? By way of example, it is worth returning briefly to our discussion of Moore's problem to see what kind of light Hintikka's formalization shed on that case. In *Knowledge and Belief*, he was able to prove that statements of the sort "*p* and I do not believe *p*" are *perfect absurdities* not because they run afoul of some kind of epistemic intuition, but because, when properly analyzed, they generate a contradiction. More importantly, the analysis allows us to recognize which epistemic commitments are involved in generating the contradiction. These commitments are formulated as rules for epistemic alternatives in model systems. So, for example, the proof of the absurdity of "*p* and I do not believe *p*" (Hintikka 1962: 68) relies on the conditions governing the semantics of sentences concerning belief. The difference between the kind of reasoning we find in Moore and Hintikka with respect to "*p* and I do not believe *p*" boils down to difference with respect to the degree of explicitness and control that the philosophers aspire to in their arguments. For Hintikka, unlike Moore, the point is to achieve the same level of explicitness in epistemology as is found in logic:

> The word "logic" which occurs in the subtitle of this work is to be taken seriously. My first aim is to formulate and to defend explicit criteria of consistency for certain sets of statements—criteria which, it is hoped, will be comparable with the criteria of consistency studied in the established branches of logic.
> (1962: 3)

Logical Omniscience and Idealized Epistemic Agents

Epistemic logic inevitably traffics in idealizations. As discussed below, the problem of logical omniscience (a product of accepting the axiom of deductive cogency or axiom K and standard possible world semantics) encouraged theorists to craft formal systems which more adequately reflected the actual properties of epistemic agents. Since real epistemic agents modify their beliefs and engage in inquiry, there was some philosophical interest in attempting to formally capture the dynamical features of inquiry. Developments since *Knowledge and Belief*, principally those since Kutchera (1976) and Lenzen (1978), attempted to integrate broader insights from modal logic with epistemic logic and have made it possible to formally model some prominent features of the dynamical nature of epistemic agency. Gärdenfors' (1988) account of belief revision was particularly important in setting the stage for a slew of dynamical models of knowledge.

Logical omniscience is related to closure properties. Axiom K can, under certain circumstances, be generalized to a closure property for an agent's knowledge which is implausibly strong: Whenever an agent *c* knows all of the formulas in a set Γ and *A* follows logically from Γ, then *c* also knows A. In particular, *c* knows all theorems (letting $\Gamma = \emptyset$), and he knows all logical consequences of any formula which he knows (letting Γ consist of a single formula).

In response to the threat of logical omniscience, some epistemologists raised the question of whether the very idea of a logic of knowledge makes any epistemological sense. For instance, Hocutt challenged the applicability of logic to any realistic account of knowledge (1972). Because there is no guarantee that a knower will recognize that it is committed to some proposition that is logically equivalent to some proposition to which it readily assents, the very idea of an epistemic logic is on slippery ground.

Some of the first proposals for solving the problem of logical omniscience introduce semantical entities which explain why the agent appears to be, but in fact is not really logically omniscient. These entities were called 'impossible possible worlds' by Hintikka (1975). The basic idea is that an agent might mistakenly count among the worlds consistent with his or her knowledge, some worlds containing logical contradictions. The mistake is simply a product of limited resources; the agent might not be in a position to detect the contradiction and might erroneously count them as genuine possibilities.

'Seemingly possible' worlds are introduced by Veikko Rantala (1975) in his urn-model analysis of logical omniscience. Rantala devised a way of alleviating the mismatch between our model theoretic reasoning about knowledge and our proof theoretic commitments: He asks us to conceive of our epistemic relationship with the world by analogy with an urn from which we can draw balls (individual units of information) one by one over time. With each new piece of information drawn from the urn, we can modify our models. The idea is, simply, that inquiry is a dynamical process in which our model of the world changes with new information. Rantala has provided a formalism which incorporates an intuitively reasonable notion of change in a model. Such change can be understood as a change in the properties of individuals of the model or a change in its universe of discourse.

Representing how the agent's model might dynamically update is one way of thinking about epistemic agency in a more realistic manner. However, on any realistic account of epistemic agency, the agent is likely to consider (albeit inadvertently) worlds in which the classical laws of logic do not hold. In this context, the general problem of establishing a set of epistemic principles for a realistic agent is unavoidable. Rantala's approach provides a way of making the appearance of logical omniscience less threatening, but at the cost of introducing a degree of arbitrariness along with impossible or seemingly possible worlds (see Rantala 1982). In Rantala's discussion of the semantics for impossible worlds (1982) the truth condition is completely free, insofar as any contradiction among an agent's beliefs can be represented by a model containing an impossible world. While logical omniscience is avoided, the price we pay is high, since no real epistemic principles hold broadly enough to encompass impossible and seemingly possible worlds (see Meyer and van der Hoek 1995: 87–88).

Some conditions must be applied to epistemic models such that they cohere with epistemic principles. Computer scientists have proposed that what is being modeled in epistemic logic is not knowledge simpliciter but a related concept which is immune to logical omniscience. The epistemic operator $K_c A$ should be read as 'agent c knows implicitly A,' 'A follows from c's knowledge,' 'A is agent c's possible knowledge,' etc. Propositional attitudes like these should replace the usual 'agent c knows A'. While there exists some variation, the locutions all suggest modeling implicit knowledge or what is implicitly represented in an agent's information state rather than explicit knowledge

(Fagin et al. 1995, and others). The agents neither have to compute knowledge nor can they be held responsible for answering queries based on their knowledge under the implicit understanding of knowledge. Logical omniscience is an epistemological condition for implicit knowledge, but the agent might actually fail to realize this condition.

There are a variety of ways of responding to these kinds of challenges. One rather unpromising approach is to deny that epistemic logic is under any obligation to connect with more general epistemological concerns (see, for example, Lenzen 1978: 34). Rather than treating epistemic logic as a purely formal exercise, a preferable response involves maintaining that epistemic logic does carry epistemological significance but in an inevitably idealized sort of way. One restricts attention to a class of rational agents where rationality is defined by certain postulates. Thus, agents have to satisfy at least some minimal conditions to simply qualify as rational. This is, for example, what Lemmon originally suggests (Lemmon 1959). One such condition would involve assuming that rational agents should acknowledge the laws of logic. For instance, if the agent knows p and $p \rightarrow q$, it should be able to recognize that q follows validly.

These 'rationality postulates' for knowledge exhibit a striking similarity to the laws of modal and epistemic logic. One can, in turn, legitimately attempt to interpret the necessity operator in alethic axioms as a knowledge operator and then justify the modal axioms as axioms of knowledge. While Lemmon constructs the rational epistemic agent directly from the axiomatization of the logic, another way of justifying the epistemic axioms involves reference to their semantical features. This is the line of thought that Hintikka pursued in *Knowledge and Belief.* Hintikka stipulated that the axioms or principles of epistemic logic are conditions descriptive of a special kind of general (strong) rationality. The statements that can be proved false by application of the epistemic axioms are not inconsistent, meaning that their truth is logically impossible. They are, rather, rationally 'indefensible.' Indefensibility is fleshed out as the agent's epistemic laziness, sloppiness or perhaps cognitive incapacity whenever to realize the implications of what he in fact knows. Defensibility, then, means not falling victim of 'epistemic negligence' as Chisholm calls it (Chisholm 1963, 1977). The notion of indefensibility gives away the status of the epistemic axioms and logics. Some epistemic statement for which its negation is indefensible is called 'self-sustaining.' The notion of self-sustenance actually corresponds to the meta-logical concept of validity. Corresponding to a self-sustaining statement is a logically valid statement. But this will again be a statement which is rationally indefensible to deny. So, in conclusion, epistemic axioms can be understood to be descriptions of rationality. This argument is spelled out in detail by Hilpinen (2002).

Common Knowledge and Distributed Knowledge

So far, this essay has discussed the epistemic properties of individual agents. However, many recent developments in epistemic logic concern the study of the formal properties of systems of interacting agents. This section introduces two of the most prominent notions in the study of multi-agent systems: common and distributed knowledge.

When we consider agents who are connected via some network we can study the effect of new information, presented to part of the (or made public to the whole) group. Formal grasp of the role of announcements in a complex network of agents has important practical consequences for our understanding of cooperation and competition. The manner in which new information moves through a multi-agent system and how it