

Idealization, epistemic logic, and epistemology

Audrey Yap

Received: 28 February 2013 / Accepted: 14 March 2014 / Published online: 10 April 2014
© Springer Science+Business Media Dordrecht 2014

Abstract Many criticisms of epistemic logic have centered around its use of devices such as idealized knowers with logical omniscience and perfect self-knowledge. One possible response to such criticisms is to say that these idealizations are normative devices, and that epistemic logic tells us how agents ought to behave. This paper will take a different approach, treating epistemic logic as descriptive, and drawing the analogy between its formal models and idealized scientific models on that basis. Treating it as descriptive matches the way in which some philosophers, including one of its founders, Jaako Hintikka, have thought about epistemic logic early in its history. Further, the analogy between the two fields will give us a way to defuse criticisms that see epistemic logic as unrealistic. For example, criticizing models of epistemic logic in which agents know all propositional tautologies as being unrealistic would be like criticizing frictionless planes in physics for being unrealistic. Each one would certainly be an unsuitable model for studying some kinds of phenomena, but is entirely appropriate for others. After outlining the analogy between epistemic and scientific models, we will discuss some ways in which idealizations are used by different research programs in epistemic logic.

Keywords Epistemic Logic · Epistemology · Idealization · Logic · Philosophy of logic

Many criticisms of epistemic logic have centered around its use of devices such as idealized knowers with logical omniscience and perfect self-knowledge. The agents of epistemic logic clearly have capacities that are not possessed by ordinary knowing agents (such as human beings). This has led some to argue against the use of epistemic

A. Yap (✉)
Department of Philosophy, University of Victoria, P.O. Box 3045,
Victoria, BC V8W3P4, Canada
e-mail: ayap@uvic.ca

logic on the basis of its failure to represent features of actual knowers accurately. One possible line of defense against these criticisms is to say that these idealizations are normative devices, and that epistemic logic tells us how agents ought to behave, rather than how they really do behave. Then it would tell us what conclusions *ought* to be drawn from existing beliefs, regardless of whether any agent with those beliefs actually does so. For instance, a recent paper by Mark Colyvan discusses idealization in several different types of theories such as decision theory and logic, treating such theories and many of the idealizations they employ as normative (Colyvan 2012). The present paper will take a different approach, treating epistemic logic as descriptive, and drawing an analogy between its formal models and idealized scientific models on that basis. I will not argue against a normative view of epistemic logic, but this paper will take for granted that it is in many ways descriptive. Further, treating it as descriptive matches the way in which some philosophers, including one of its founders, Jaako Hintikka, have thought about epistemic logic early in its history.¹ The analogy between the two fields will give us a way (different from Colyvan's) to defuse criticisms of epistemic logic that see it as being unrealistic. For example, criticizing models of epistemic logic in which agents know all propositional tautologies as being unrealistic would be like criticizing frictionless planes in physics for being unrealistic. Each one would certainly be an unsuitable model for studying *some* kinds of phenomena, but is entirely appropriate for others. This paper will not question the claim that many properties of represented agents are unrealistic idealizations of real-world agents, and that there are many ways in which the K operator of epistemic logic does not properly track the real-world concept of knowledge. What will instead be argued is that such a criticism of epistemic logic misses the point.

1 Epistemic logic and its critics

Epistemic logic was developed by logicians in works such as (Hintikka 1962), but its connection to epistemology itself has not always been clear. More recent works such as (van Benthem 2006; de Bruin 2008; Hendricks and Symons 2006) have discussed connections between the two, and go some way to addressing criticisms such as (Hocutt 1972) and (Girle 1998). But while these works make a strong case that epistemic logic has a lot of usefulness for epistemology, they do not entirely answer the worries raised that it fails accurately to represent real epistemic agents. (Gochet and Gribomont 2006) give a more complete history of epistemic logic and its problems (such as logical omniscience) that we discuss here. The problem of logical omniscience specifically is also discussed in (Sim 1997). In this section of the paper, I will briefly present the basics of contemporary epistemic logic as well as an argument by Max Hocutt that questions the very possibility of an epistemic logic.

There are many varieties of contemporary epistemic logic, but it suffices for our discussion to present a simple propositional version that models a single agent's knowl-

¹ And there really does seem to be space to treat epistemic logic as a discipline with both normative and descriptive components. But since the concerns in this paper have principally to do with the accuracy of epistemic logic's depiction of agents, we will not discuss the normativity issue in any detail.

edge. The way in which it does so is by using Kripke models, a standard device of classical modal logic. A Kripke model in modal logic is a tuple $M = \langle W, R, V \rangle$, in which W is a set, interpreted as a set of possible worlds, R is a binary relation between elements of W , and V is a function taking each proposition letter to a subset of W , interpreted as the set of worlds in the model at which that proposition is true. A logic of knowledge often interprets the R relation to represent indistinguishability to the agent, typically by ensuring that the relation is reflexive, symmetric, and transitive (giving us the modal logic S5). A logic of belief drops the reflexivity requirement (giving us the modal logic KD45), since while knowledge ought to be veridical, beliefs need not be. If we want to remain more neutral on the substantive assumptions underlying the R relation, we can refer to it as an accessibility relation. The language of modal logic adds modal operators such as K_a and B_a to basic propositional logic, where $K_a\varphi$ and $B_a\varphi$ can be read as “ a knows that φ ” and “ a believes that φ ,” respectively. Each is defined in terms of truth in all accessible worlds.

Now, while Hintikka’s own system of epistemic logic differs from that used by many contemporary epistemic logicians, these contemporary systems can still be traced back to his work. And the worry that we will outline applies to the discipline generally, being largely independent of which specific version of the logic we use. Indeed, the first sentence of (Hintikka 1962) invites a certain kind of worry when it is claimed that “The word ‘logic’ which occurs in the subtitle of this work is to be taken seriously” (Hintikka 1962, p. 3). The worry has to do with the sense in which we can expect there to be such a thing as a logic of knowledge, and the way in which we can connect epistemic logic to epistemology.

Some contemporary authors have tried to connect epistemic logic and epistemology in terms of the former’s usefulness to the latter. Both (van Benthem 2006) and (Hendricks and Symons 2006) outline several developments in modern epistemic logic such as the addition of logical dynamics and epistemic logics for multi-agent systems, arguing that they parallel issues in epistemology proper. Each paper treats epistemic logic as a discipline that can serve to clarify philosophical concepts and debates related to knowledge and belief. (de Bruin 2008) also considers some applications of epistemic logic to historical problems in epistemology, such as a formalization of Cartesian skeptical doubt. He also briefly considers the question of the connection between the actual verb “to know” and the idealized way in which knowledge is treated in the formalism of epistemic logic—a question at the heart of Hocutt’s criticism of epistemic logic. de Bruin addresses Hocutt’s work and proposes some technical responses, but says little philosophically to address Hocutt’s central criticism.

The criticism can be phrased as a dilemma: either epistemic logic fails to be *epistemic* or it fails to be *logic*. This dilemma, Hocutt believes, arises as a result of several counterexamples to theorems of epistemic logic. Responses to these counterexamples will either render the term ‘know’ vacuous—thereby failing to connect to epistemology, or deny that theorems of epistemic logic need to be true—thereby failing to remain a logic (Hocutt 1972, p. 436). We will focus here on the criticism that the knowledge captured by epistemic logic fails to connect to real-world knowledge.² For example,

² More contemporary authors such as (Girle 1998) have raised the problem of epistemic logic’s unrealistic features, even accounting for developments in the discipline since (Hintikka 1962). So while our focus will

a basic principle of epistemic logic is its version of the K axiom of modal logic:

$$K_a(\varphi \rightarrow \psi) \rightarrow (K_a\varphi \rightarrow K_a\psi).$$

In other words, if a knows that φ implies ψ , then his knowing that φ implies that he also knows ψ . This means that a 's beliefs are closed under the application of modus ponens. The counterexample Hocutt gives is that of the Logically Obtuse Man (LOM), who does not always deduce the consequences of his beliefs (Hocutt 1972, p. 435). And indeed, it seems that most of us fail to deduce *all* the consequences of our beliefs. While in simple propositional cases, the K axiom may seem realistic, for more complex instances of φ and ψ , many of us are likely logically obtuse. We will follow Hocutt in considering a possible response to this counterexample, attributed to (Lemmon and Henderson 1959): logically perspicuous knowers (LPKs).

An LPK is one who, unlike the LOM, always deduces the consequences of his beliefs. While Lemmon admits that normal people are not LPKs, he writes that we may treat an agent X of epistemic logic as

a kind of logical fiction, the rational man. Then, if X knows that if . . . then — and further knows that . . . , being rational he knows that . . . , so that [the K axiom] is satisfied. (A rational man knows (at least implicitly) the logical consequences of what he knows.) (Lemmon and Henderson 1959, p. 39)

Then in order to circumvent the counterexample presented above, we can simply say that epistemic logic is not the logic of the LOM, but rather the logic of the LPK. We can look at this response in two ways. One is to say that the sense of “know” when we say that an LOM knows that φ is different from the sense of “know” when we say that an LPK knows that φ . And it is the latter sense of the word that is captured by the K operator of epistemic logic. This denies the applicability of the counterexample, but disconnects epistemic logic from actual knowledge. For if epistemic logic is the logic of idealized knowledge instead of ordinary knowledge, we find ourselves with the first horn of Hocutt's dilemma. Lemmon's response to this concern is to admit the disconnect between the ordinary sense of knowledge and the sense in which the LPK knows the logical consequences of his beliefs; but rather than considering this to be a failing of epistemic logic, he writes that epistemic logic may play a normative role in deciding what counts as knowledge. In that case, what the LPK knows is what we ought to know, though we might not (Lemmon and Henderson 1959, p. 40).

A second way of seeing the response to the counterexample is to say that the sense of “knowledge” is in fact the same, and is in fact ordinary knowledge. The difference is that agents such as the LPK have different capacities, and will know more facts because of their greater deductive abilities. So it is not the sense of “knowledge” that changes, but the epistemic agents we are considering. Unfortunately, this way of viewing the response still leaves us with the first horn of Hocutt's dilemma. Even though it still allows epistemic logic to be the logic of knowledge, it is not the logic of ordinary

Footnote 2 continued

be on Hocutt's critique, he is certainly not the only one to have made the point that features such as logical and deductive omniscience are deeply unrealistic when we consider actual knowers.

agents, but of idealized agents. This also has the effect of separating epistemic logic from epistemology, as the study of ordinary agents' knowledge.

One potential way out of this dilemma that Hocutt discusses, but ultimately finds flawed, is the idea that epistemic logic might have a normative interpretation.³ Among the flaws he cites is the idea that such an interpretation does nothing to clarify the concepts under consideration. As a related worry, he finds it to reverse the explanatory order of things: while a normative interpretation would say that a contradiction is illogical because it would be illogical to think one, we should instead say that we should not think a contradiction because it is illogical (Hocutt 1972, p. 451). Yet others such as (McLane 1979) find a normative interpretation of the idealized nature of epistemic logic more appealing. The LPK to which Hocutt objects could easily be interpreted as the kind of agent that we should aspire to emulate in our beliefs.

However, many defenders of epistemic logic do not view the discipline as normative. Hintikka admits that epistemic logic only tells us something definite about truth and falsity in a world in which everyone is logically omniscient (Hintikka 1962, p. 36); nevertheless, he does not say that we *ought* to try to approximate this world. Rather, with respect to the truths of epistemic logic, he writes that

Logical truths are not truths which logic forces on us; they are not necessary truths in the sense of being unavoidable. They are not truths we *must* know, but truths which we *can* know without making use of any factual information. . . . The fact that the so-called laws of logic are not “laws of thought” in the sense of natural laws seems to be generally admitted nowadays. Yet the laws of logic are not laws of thought in the sense of commands, either, except perhaps laws of the sharpest possible thought. Given a number of premises, logic does not tell us what conclusions we ought to draw from them; it merely tells us what conclusions we may draw from them—if we wish and we are clever enough. (Hintikka 1962, p. 37)

This seems to rule out a normative interpretation of epistemic logic in which epistemic logic tells us which conclusions we ought to draw from our beliefs. Instead, it tells us which conclusions we could draw without additional factual information, or which conclusions an LPK would draw from their beliefs. (de Bruin 2008) also argues explicitly against a normative interpretation of epistemic logic. For one, it might not even make sense to talk about our obligation to form particular beliefs: “one may even doubt whether one can have obligations to form beliefs at all, since, as one may argue, the formation of beliefs would be involuntary.” (de Bruin 2008, p. 124)⁴ And there could be further problems if we extend this to conclusions about what we ought to know. A normative interpretation of what we ought to know might oblige us to render our beliefs true, even going beyond the obligation to believe them. This, however,

³ While Hocutt writes that no epistemic logician has suggested such a proposal, we have already seen (Lemmon and Henderson 1959) talking much along these lines.

⁴ This is admittedly a complicated claim. (Feldman 1988) considers the issue of doxastic voluntarism, arguing that it does make sense to talk about obligations that we may be incapable of fulfilling. This would mean that even if belief formation is involuntary, we may still have obligations to believe. I will not treat this issue further here, however, instead taking it for granted that a largely descriptive interpretation of epistemic logic is desirable.

leaves us with the first horn of Hocutt's initial dilemma, which is that epistemic logic no longer seems to be a logic of knowledge as we ordinarily understand it. Instead, it is a logic of the knowledge of ideal agents.

In what follows, we will give a non-normative interpretation of epistemic logic without the circularity that Hocutt criticizes. There is no question that epistemic logic uses many idealizations that are not reflected in actual agents, and I will grant Hintikka and de Bruin that these are not normative constraints on actual agents. I will argue that there is an analogy between the use of these idealizations in logic and the uses of idealizations in natural science, which are generally not seen as normative either. While the best way to think of these idealizations may yet be controversial (see for instance (Frigg 2010; Godfrey-Smith 2009)), the analogy allows us to tie their value to their usefulness. In this way, we can take for granted that the assumptions of systems of epistemic logic are unrealistic, yet still treat such systems as useful for the analysis of real-world knowledge.

2 Idealization in science

In order to draw the analogy between models in epistemic logic and models in science, we will first discuss the nature of idealization in science, and consider the extent to which epistemic logic's models can be seen as idealizations in that sense. (Weisberg 2007) outlines three kinds of scientific idealization: *Galilean*, *minimalist*, and *multiple-models*. An important point he makes in drawing this distinction, however, is that idealization is an activity—and the distinctions between different types of idealizations are based in the types of activity involved and the ways in which these activities are justified.

Galilean idealization distorts theories describing complex systems in order to make them more computationally tractable. This is typically given a pragmatic justification, and de-idealization generally takes place when additional computational power is made available. For example, the calculations of molecular properties in computational chemistry might use Galilean idealization. In these cases, the results are acknowledged as approximations, though these approximations become increasingly accurate as computers become more powerful.

Minimalist idealization constructs theoretical models that only include the most salient factors of the phenomenon in question. These minimal models are typically developed to capture important causal relationships, generally for the purposes of explanation. For example, in explaining Boyle's law, theorists will (falsely) assume that molecules do not collide, because the collisions make no difference to the phenomenon. While these minimal models may coincide with the models produced by Galilean idealization, the difference lies in the way in which the activity is justified. Galilean idealization has a pragmatic justification, while minimalist idealization is said to capture what "really matters" about the phenomenon. As such, the distortions introduced by Galilean idealization will decrease as computational power renders problems more tractable; but core features of a phenomenon required for explanation will not change in the same way. This difference in justification means that, even if the models produced by these different methods do coincide at some point in the

course of a science, we should not expect them to continue to coincide as the science progresses.

Finally, multiple-models idealization involves the construction of different and incompatible models of a phenomenon, each of which represents the causal structure in a different way. Yet there is no strong expectation of arriving at a single best model that combines all of these. Justification for this type of idealization can vary. It might be based on the idea that different scientists have different goals—as such, one research program might favor simplicity over accuracy, while another might require the reverse. It might also parallel justification for minimalist idealization, but note that where the phenomenon is particularly complex, multiple models might be required to capture the different causal relationships. For example, the United States National Weather Service employs three distinct complex models to represent global weather. In this case, the approach is justified by its predictive power.

Now, while knowledge is generally not something that epistemic logicians model for the sake of prediction, multiple-models idealization seems to be a natural way of describing the activity. After all, there is a definite plurality of systems of epistemic logic, each of which makes different idealizing assumptions for the sake of representing slightly different aspects of agents' knowledge. We can also see different researchers thinking of epistemic logic in terms of Galilean idealization, working to develop more general systems that eliminate different idealizing assumptions in order more accurately to model agents' behavior. Finally, we can see some aspects of minimalist idealization in the choices of which realistic features of agents will be represented in models, and which will be ignored. As such, this paper will draw an analogy between idealization in logic and idealization in the sciences in terms of the choices and tradeoffs made by model-builders, following the framework from (Weisberg 2007). In order to discuss these choices, we will first introduce the idea of a *representational ideal*. These are

the goals governing the construction, analysis, and evaluation of theoretical models. They regulate which factors are to be included in models, set up the standards theorists use to evaluate their models, and guide the direction of theoretical inquiry. Representational ideals can be thought of as having two components: *inclusion rules* and *fidelity rules*. Inclusion rules tell the theorist which kinds of properties of the phenomenon of interest, or *target system*, must be included in the model, while fidelity rules concern the degrees of precision and accuracy with which each part of the model is to be judged. (Weisberg 2007, p. 648–9)

The most relevant representational ideals for our purposes are simplicity, 1-causal, and p-general.⁵ While these are described in terms of the natural sciences, we can describe them in terms of the construction of formal models more generally, by dropping the causal language. After all, while we are seeing both types of program as attempting to represent complex phenomena, accounts in the sciences tend to focus more on causal factors than accounts in logic do. But whether the factors involved are

⁵ (Weisberg 2007) lists other representational ideals, but these have been omitted as being less relevant to the type of modeling in epistemic logic.

explicitly causal or not should not make a significant difference to the analogy being drawn here.

The first representational ideal, simplicity, dictates that the model should include as little as possible while remaining consistent with the existing fidelity rules. A simpler model might allow us to focus on a property of interest, even if we miss out on representing the interaction between that property and others. In contrast, 1-causal ensures that our models still include all of the key factors involved. While its inclusion criteria only dictate that the models include the core factors, its fidelity criteria dictate that *all* such factors, whatever they may be relative to the phenomenon in question, are represented. As such, the specific factors required to model any particular phenomenon will almost certainly vary relative to our goals. Finally, we want our models to capture as many possible instances of the phenomenon in question, which is the representational ideal of p-generality. So the three ideals that we might see employed are the models' being simple, including all relevant factors, and being as general as possible.

Now that we have outlined these three representational ideals, the next section will look at several different works of epistemic logic through the lens of Weisberg's framework, considering how they employ the representational ideals discussed here. However, before turning to that discussion, a few disclaimers will be needed. Just as there is no single uniform account of the goals of idealization in science, we should not expect a uniform account of idealization in epistemic logic. The purpose of this paper is to argue for an analogy between the use of idealization in some important research programs in epistemic logic, and some uses of idealization in the sciences. Also, we are effectively grasping the second horn of Hocutt's dilemma, by denying that theorems of epistemic logic need to be true. Does this lead to the conclusion that epistemic logic is no longer a logic? That depends. Hocutt could not be claiming that epistemic logic is not a formal system, since in that latter sense it is surely a logic. On the other hand, the implicit multiple-models stance we have taken supposes that there is no single system of epistemic logic that will accurately represent every aspect of knowledge. This holds even if we think that knowledge is a single coherent phenomenon—what the discussion of scientific idealization tells us is that any given model may be insufficient to represent a sufficiently complex phenomenon. So under the assumption of logical monism, epistemic logic would likely not be a logic. But to what extent do we want to be logical monists about epistemic logic? A monist philosophical view is rather out of step with the discipline, in which it is more common to construct new systems to solve problems and represent different aspects of knowledge than it is to debate which system might be the single correct one. Yet a pluralistic multiple-models perspective does not necessarily commit us to anti-realism either (Pincock 2011). The connected issues of realism, monism, and pluralism will not be discussed further in this paper, however. The concern in the remaining section will simply be to show that idealization analogous to that in the sciences takes place in epistemic logic.

3 Scientific models, epistemic models

In this section, we will survey some different research programs in contemporary epistemic logic in order to determine the nature of their use of idealization. The underlying contention will be that each one of these programs attempts to capture

something about the nature of knowledge that might not yet be satisfactorily modeled by existing formal systems. A relatively common method of motivating the introduction of a new formal system is to point out a real life situation that is not properly represented by existing systems, but whose salient features the new system can adequately model. Such formal systems are not attempting to model knowledge in its entirety, but will generally extend a basic epistemic logic in order to make it more realistic in a specific way. In this sense, we can see that the choice of which idealizations each system eliminates, and which it retains depend on its particular modeling goals.

It is also worth noting that the issues we have outlined regarding idealization are not merely problems with epistemic logic, but can be said to apply to knowledge representation generally. As such, after demonstrating a few ways in which research programs in epistemic logic engage with issues about idealization, we will turn to some uses of idealization in other approaches to the formalization of knowledge.⁶

3.1 Resource-bounded agents

Idealized knowers of epistemic logic typically have no restriction on their memories. Indeed, a common idealizing assumption in many systems is that of *perfect recall*, which in effect ensures that agents remember all that has previously happened to them. This means that they remember the sequence of events that has lead to the present moment, as well as their previous informational states. This is one unrealistic feature of ideal agents to which several alternatives have been considered. For instance, (van Benthem and Liu 2004; Liu 2009) consider several different versions of agents with bounded memories, including memory-free agents (agents who only remember the last event). This de-idealization takes place in systems of Dynamic Epistemic Logic, in which operations on epistemic models represent the informational effects of events taking place over time. When these systems idealize by incorporating perfect recall, agents' informational states are in some sense cumulative, meaning that they never lose information they had previously gained.

While (Liu 2009) does take epistemic logic to be a normative system, she also notes that its value depends on its having a connection with reality, and that actual reasoning does take place among communities of diverse agents. Her concern with the many different types of agents there might be is an instance of p-generality, the ideal of representing as many different instances as possible of the phenomenon in question. For example, she distinguishes five different types of agents in belief revision, based on the extent to which observations affect their informational states (Liu 2009, p. 39–40). This distinction comes in addition to the development of models representing memory-bounded agents. We can also see in her discussion of future research the description of a program that employs Galilean idealization:

⁶ The models of epistemic logic we consider are also all research programs in the *external* approach to knowledge representation. That is, we consider modeling agents from the point of view of an external observer. An *internal* approach to epistemic logic from the point of view of one of the agents in the situation is given in (Aucher 2010). However, autoepistemic logic is an alternative knowledge representation system we outline that can be seen as an instance of the internal approach.

there remains the issue whether one can have a general view of the natural “parameters” that determine differences in behavior of logical agents. Our analysis does not provide such a general account, but at least, it shows more richness and uniformity than earlier ones. Second, even with all these parameters on the map, we have not yet found one framework for all these sources. In particular, current work on limitations of inferential or computational powers should be integrated with that one observational and learning diversity. Our next ambition would be to put all these features together in one plausible computational model of an agent as an information-processing and decision-making device, with modules for perception, memory, and inference which can communicate and share information (Liu 2009, p. 51)

It is clear that the intention here is not to prescribe proper epistemic behavior for agents, but to provide a formal model for representing their actual behavior. We also find features explicitly listed for inclusion in the desired general model, reflecting the salient features of the phenomenon in question. In this particular piece of work, the interest is in agents with different memory capacities, so we de-idealize by removing the requirement of perfect recall. But while the goal is to provide more plausible models of agents in the world, features of these agents that have nothing to do with their ability to process information and make decisions on that basis are irrelevant to the model. For example, there are no attempts in this work to remove the idealizations of agents’ unlimited logical abilities. We will turn next to a system that does maintain the assumption of perfect recall, but represents more realistic agents in a different way.

3.2 Implicit and explicit knowledge

One of the issues that has been raised in our discussion of epistemic logic’s unrealistic idealizations is that of logical omniscience, or agents’ ability to deduce all the logical consequences of their beliefs. One way to represent more realistic agents is to draw a formal distinction between *implicit* and *explicit* knowledge. Implicit knowledge is essentially the knowledge captured by basic epistemic logic, which is the knowledge that an agent could possess if she were able to close her knowledge under logical consequence. We could think of an agent’s implicit knowledge as the set of things that she would be capable of knowing given unlimited cognitive resources. Explicit knowledge, on the other hand, is the more ordinary and resource-limited sense of knowledge. What an agent knows explicitly are the things that she is aware of, and that she has in fact deduced based on the things she explicitly knows. This means that fidelity criteria will apply to explicit knowledge, not implicit knowledge. In order to formalize this distinction, we need to complicate the existing framework of epistemic logic by adding new mechanisms. For example, (Velazquez-Quesada 2009) extends the framework of dynamic epistemic logic with a new type of update that represents an agent’s drawing an inference from her existing knowledge base. Again, this modifies the mechanism that modifies epistemic models to reflect the occurrence of informational events.

In this extended framework, an agent’s explicit information at a world is represented by a set of propositions. Unlike her implicit knowledge, this will normally not be closed under logical consequence. We represent rules that an agent could apply to her belief

set as pairs, where the first element is a finite set of premises, and the second element is a formula in the language that we interpret as the conclusion. For instance, we could have a rule that is an instance of modus ponens represented as the pair $(\{p, p \rightarrow q\}, q)$. Agents might have different sets of rules that they can explicitly apply at any given world, just as they might have different information that they are explicitly aware of at any given world. The framework is also sufficiently flexible that rules are not restricted to truth-preserving inference patterns. The rules an agent can access might also change as the model becomes updated.

In ordinary Dynamic Epistemic Logic, agents can learn the truth-value of a proposition, for instance, that p is false, and this updates their informational state accordingly. In a logic of explicit and implicit information, an agent could become informed about a new rule (as in a situation in which he learns about a new inference pattern he could apply). He might also deduce new rules based on his existing rule set, just as he might use his rules to deduce consequences of his explicit information, all of which provides a finer-grained representation of agents' knowledge than that of standard epistemic logic. This flexibility, allowing different rules to be applied in different situations, is an instance of p -generality, in which we try to allow for the representation of as many possible instances as we can of the phenomenon in question. In this case, we are interested in an agent's logical ability. Other features of more realistic situations are ignored as being relatively irrelevant to this particular issue. For example, this framework does not build in issues of memory and forgetting previously known propositions. Regardless, this use of systems that distinguish between implicit and explicit knowledge is another contribution to the representation of resource-bounded agents. In these systems, agents are still capable of deducing the logical consequences of their beliefs, but unlike in ordinary epistemic logic, we do not see them as already having closed their information under logical consequence. As such, it de-idealizes in allowing for both a more faithful representation of agents, as well as allowing us to represent a wider variety of agents, in accordance with considerations of fidelity and generality.

3.3 Kripke models and gettier problems

Most recently, an explicit application of epistemic logic to a philosophical problem is the treatment of Gettier problems in (Williamson 2012). These cases are intended to be counterexamples to the view that justified true belief is sufficient for knowledge. Williamson's goal in this paper is to provide a class of epistemic models in which agents have justified true belief without knowledge, thereby showing that the phenomenon is formalizable.

Much of Williamson's technical machinery is the familiar machinery of epistemic logic. Kripke models still form the basis of the system, in which we have sets of possible worlds and accessibility relations between those worlds, representing the agent's uncertainties. The modification that he makes is to represent worlds in a more fine-grained fashion. Rather than tracking only the propositions true and false at each world, we represent the fact that some features of the world might better be modeled as parameters (for instance, temperature). Worlds are ordered pairs $\langle e, f \rangle$ in which the first component represents the actual value of the parameter, while the second represents its apparent value for the agent in question. In each world, then, there is an available mea-

sure of the gap between appearance and reality. Accessibility relations for epistemic agents are set such that indistinguishable worlds must appear the same to them.

Despite this modification, many existing features of standard epistemic logic remain. For instance, formulas are still interpreted according to the usual semantics. In particular, an agent knows that p is true iff p is true in every world accessible for that agent. This, as we have already seen, gives rise to logical omniscience, or to agents like Hocutt's LPK. Williamson is explicit about this fact, but notes that the idealization involved is benign, or perhaps even beneficial when we consider the goals of the system. After all, if even the LPK lacks knowledge in a Gettier case, this only makes the phenomenon seem more robust (Williamson 2012, p. 6). Another idealization that Williamson believes to be false of actual agents, having argued against it in (Williamson 2000), is the transparency of appearances to agents. But just as with logical omniscience, if even agents that are idealized in this sense lack knowledge in Gettier cases, we should not expect ordinary non-idealized agents to have knowledge in those situations either (Williamson 2012, p. 7). In Williamson's models, then, the idealization improves, rather than hinders, the analysis. The fact that the represented agents are unrealistic is not a source of concern; the goal is not to represent human agents accurately, but to model a particular epistemic phenomenon to which human agents are subject.

We may however, have a worry with respect to the generality of Williamson's models. While he does give some examples of Gettier problems that can be modeled using his formalism, there are other types of Gettier problem that could potentially be more problematic. Williamson's cases are more like "fake barn" cases, in which an agent lacks relevant beliefs, than they are like Gettier cases in which agents derive justified true beliefs from justified false beliefs (Williamson 2012, p. 19). So while Williamson claims that some of his models represent the more traditional type of Gettier model, there may nevertheless be a concern that the framework be able to represent sufficiently many instances of the phenomena being represented, thereby satisfying p -generality.

3.4 Autoepistemic logic

Part of the study of artificial intelligence in computer science is the representation of agents' knowledge. *Autoepistemic logic* is a contribution to artificial intelligence that was developed as a response to certain methodological problems with the use of existing nonmonotonic logics to represent agents' beliefs (Moore 1985).⁷ Nonmonotonic logics themselves can be seen as ways of de-idealizing classical logics, particularly insofar as logic is used to represent reasoning. There are certainly cases in which we will retract a previously held and well-justified belief in light of further evidence, so the set of a real agent's beliefs does not increase monotonically. The standard example of such reasoning is as follows, and demonstrates the need for nonmonotonic formal systems:

⁷ Though Moore himself attributes the term 'autoepistemic' to an unpublished note by Stalnaker.

If we know that Tweety is a bird, we will normally assume, in the absence of evidence to the contrary, that Tweety can fly. If, however, we later learn that Tweety is a penguin, we will withdraw our prior assumption. If we try to model this in a formal system, we seem to have a situation in which a theorem P is derivable from a set of axioms S , but is not derivable from some set S' that is a superset of S . The set of theorems, therefore, does not increase monotonically with the set of axioms; hence this sort of reasoning is said to be ‘nonmonotonic’. (Moore 1985, p. 75–6)

But the type of reasoning (Moore 1985) is concerned with is *autoepistemic*, or the type of reasoning an agent employs about her own beliefs. Moore adds a belief modality L to ordinary propositional logic, whose dual M is interpreted as consistency (a formula is consistent when its negation is not believed). The formal semantics resemble Hintikka’s, in that both use belief sets. An agent’s belief set in autoepistemic logic is an *autoepistemic theory* T , such that a formula $L\varphi$ holds when $\varphi \in T$, and an ordinary non-modal formula holds when it is true in the world in which the agent is situated.

An agent’s belief set is built up by means of certain closure conditions, representing the conclusions that an *ideal agent* would infer on reflecting upon her own beliefs and applying ordinary propositional logic. Moore chooses this in accordance with the representational ideal of simplicity, ideal agents being a simpler mechanism to implement than more complex agents. For instance, agents infer any tautological consequence of their existing beliefs. We then define a *stable extension* of a belief set to be one in an ideal agent would not draw any further conclusions that are not yet in the extended set. However, these closure conditions do not necessarily yield a unique extension of a given belief set, and it is possible to define belief sets which have no stable extensions. Moore discusses various relationships between his system and already defined systems of nonmonotonic logic. This speaks to the generality of his system, in that it is able to give an analysis of systems of nonmonotonic logic that explain some of their peculiar features. But for our purposes, the most important factor is what justifies his acceptance of agents’ logical perspicacity. Before defining the closure conditions, Moore poses the question

what do we want from a notion of inference for the logic? From an epistemological perspective, the problem of inference is the problem of what set of beliefs (theorems) an ideally rational agent would adopt on the basis of his initial premises (axioms). Since we are trying to model the beliefs of a rational agent, the beliefs should be sound with respect to the premises; we want a guarantee that the beliefs are true provided that the premises are true. Moreover, since we assume that the agent is *ideally* rational, the beliefs should be semantically complete; we want them to contain everything that the agent would be semantically justified in concluding from his beliefs and from the knowledge that they are his beliefs. An autoepistemic logic that meets these conditions can be viewed as a competence model of reflection upon one’s own beliefs. Like competence models generally, it assumes unbounded resources of time and memory, and is therefore not a plausible model of any finite agent. It is, however, the model

upon which the behavior of rational agents ought to converge as their time and memory resources increase. (Moore 1985, p. 81)

So Moore is quite explicit that he is interested in the behavior of ideally rational agents, and that these models are not plausible representations of ordinary knowers. Nevertheless, this idealization performs a function, namely of providing a model to which he believes behaviors will converge. This is in some ways the converse of Galilean idealization, in which the models improve by becoming more complex. Under this picture, the phenomena under consideration are the things that we can imagine improving. This is arguably a plausible way of thinking of Galilean idealization in a computer science context, in which intelligent systems are both the things being modeled and things that the researchers are creating. The idea is still that the models and phenomena will converge on each other, but where researchers in the physical sciences do not have control over the natural phenomena they study, researchers in computer science do have control over at least some of the phenomena they are studying.

3.5 Dempster-Shafer theory

The final formalism for representing knowledge and beliefs that we will consider is a probabilistic one called Dempster-Shafer Theory. This is primarily discussed in (Shafer 1976), which introduces the principal formalism. However, (Shafer 1990) discusses the theory in a more speculative way, commenting on the role Dempster-Shafer Theory plays in AI, as well as how we ought to view belief functions generally. The basic idea behind the theory is that it allows us to combine different pieces of evidence. Suppose we have a question in mind and have different pieces of evidence indicating the answer to that question. Based on our subjective beliefs about, say, the reliability of those sources of evidence, and whether those sources are independent, we can use Dempster-Shafer Theory to obtain a degree of belief about the answer to the original question. More generally, we obtain a degree of belief about one answer from degrees of belief about related answers.

Shafer's views about the place of his work in the overall discipline of knowledge representation have not remained static, however, and he discusses the changes in his position: while he originally believed it to be a maximally general tool for representing subjective judgments of probability, he eventually came to a more pluralistic point of view, in which belief functions are one among many ways of modeling such judgments. This can easily be seen as a shift to a multiple-models view. He also clearly acknowledges that his theory is an idealization, but points out several ways in which it can nevertheless be useful, describing the theory of probability as

really the theory of an ideal picture in which belief, fair price, and knowledge of the long run are bound together. Probabilities in this ideal picture are long-run frequencies, but they are also degrees of belief, because they are known, and nothing else that is relevant is known. This ideal picture seldom occurs in nature, but there are many ways of making it relevant to real problems. In some cases, the ideal picture serves as a standard of comparison... In other

cases, we simulate the ideal picture ... and then deliberately entangle this simulation in a real problem... In other cases, we draw an analogy between the state of knowledge in the ideal picture and our evidence in a real problem. (Shafer 1990, p. 2)

Different ways of using probability (belief functions included) may be better applicable than others in particular situations, but this illustrates the need for multiple models. Dempster-Shafer Theory is already fairly complex, computationally, which suggests a difficulty in generalizing it to include more features of human reasoning. So if we cannot trade off for more generality without sacrificing too much simplicity, we need to make sure we have several different theories that might each be able to cover different applications that we are interested in.

4 Conclusion

The previous examples have given a picture of some ongoing research programs in epistemic logic that showcase ways in which the field uses idealizing assumptions. While many of these projects do seek to describe the behavior of real agents, it is acknowledged that some simplifying assumptions will be required for the sake of tractability. The appropriateness of using formal models to study epistemological issues generally is a further issue, but, if this paper is right, then that appropriateness does not stand or fall with the presence of some idealizing assumptions. The question is whether, despite the inevitable idealization, the formal models can still give us insight into actual phenomena.

In that case, criticisms of epistemic logic from Hocutt and other, more contemporary, philosophers, miss the point. A paper such as (Girle 1998), for instance, that outlines several unrealistic features of formal systems of knowledge and belief does not have to be seen as a criticism. Rather, it can be seen as programmatic, outlining ways in which epistemic logics might *better* describe actual agents. Given that the field is not the static study of a single logical system, but is continually developing new systems and extending existing ones, it seems likely that all kinds of unrealistic features could be dropped—whether or not it would be practical to do so given the likely complexity of the resulting systems would be the main limiting factor. But given that this is the case with many formal models of physical systems, as we have seen in the outline of different uses of idealization in science (Weisberg 2007), the presence of idealizing assumptions is not exclusive to epistemic logic.

Indeed, what we have argued in this paper is that epistemic logic is a field that attempts to model the complex phenomena of knowledge and belief among agents, and as such is more similar than it might first seem to scientific fields that also attempt to model complex physical phenomena. In that case, criticism of a system of epistemic logic for implementing unrealistic assumptions would need some pragmatic justification explaining why those assumptions *ought* to be dropped. Since we take it for granted that some idealizing assumptions will be made in formal models, we should focus on *which* assumptions ought to be made instead of the fact of their being made in the first place.

References

- Aucher, G. (2010). An internal version of epistemic logic. *Studia Logica*, 94(1), 1–22.
- Colyvan, M. (2012). Idealisations in normative models. *Synthese*, 190, 1337–350.
- de Bruin, B. (2008). Epistemic logic and epistemology. In V. F. Hendricks & D. Pritchard (Eds.), *New Waves in Epistemology* (pp. 106–136). Basingstoke: Palgrave Macmillan.
- Feldman, R. (1988). Epistemic obligations. *Philosophical Perspectives*, 2, 235–256.
- Frigg, R. (2010). Models and fiction. *Synthese*, 172(2), 251–268.
- Girle, R. (1998). Logical fiction: Real vs. ideal. *Lecture Notes in Computer Science*, 153(1), 542–552.
- Gochet, P. & Gribomont, P. (2006). Epistemic logic. In Dov M. Gabbay & John Woods, (Eds.), *Handbook of the History of Logic*, (vol. 7, pp. 99–195). North-Holland.
- Godfrey-Smith, P. (2009). Models and fictions in science. *Philosophical Studies*, 143, 101–116.
- Hendricks, V., & Symons, J. (2006). Where's the bridge? epistemology and epistemic logic. *Philosophical Studies*, 128, 137–167.
- Hintikka, J. (1962). *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Ithaca: Cornell University Press.
- Hocutt, M. (1972). Is epistemic logic possible? *Notre Dame Journal of Formal Logic*, XII, 1 (4), 433–453.
- Lemmon, E., & Henderson, G. (1959). Is there only one correct system of modal logic? *Proceedings of the Aristotelian Society*, 33, 23–56.
- Liu, F. (2009). Diversity of agents and their interaction. *Journal of Logic, Language, and Information*, 18(1), 23–53.
- McLane, E. (1979). On the possibility of epistemic logic. *Notre Dame Journal of Formal Logic*, XX(3), 559–574.
- Moore, R. C. (1985). Semantical considerations on nonmonotonic logic. *Artificial Intelligence*, 25, 75–94.
- Pincock, C. (2011). Modeling reality. *Synthese*, 180, 19–32.
- Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton: Princeton University Press.
- Shafer, G. (1990). Perspectives on the theory and practice of belief functions. *International Journal of Approximate Reasoning*, 4, 323–362.
- Sim, K. M. (1997). Epistemic logic and logical omniscience: A survey. *International Journal of Intelligent Systems*, 12(1), 57–81.
- van Benthem, J. (2006). Epistemic logic and epistemology: The state of their affairs. *Philosophical Studies*, 128, 49–76.
- van Benthem, J., & Liu, F. (2004). Diversity of logical agents in games. *Philosophia Scientiae*, 8(2), 165–181.
- Velazquez-Quesada, F. R. (2009). Inference and update. *Synthese*, 169(2):283–300.
- Weisberg, M. (2007). Three kinds of idealization. *Journal of Philosophy*, 104(12), 639–659.
- Williamson, T. (2000). *Knowledge and Its Limits*. Oxford: Oxford University Press.
- Williamson, T. (2012). Gettier cases in epistemic logic.