Epistemic Logic and Epistemology*

Wesley H. Holliday

University of California, Berkeley

Abstract

This chapter provides a brief introduction to propositional epistemic logic and its applications to epistemology. No previous exposure to epistemic logic is assumed. Epistemic-logical topics discussed include the language and semantics of basic epistemic logic, multi-agent epistemic logic, combined epistemic-doxastic logic, and a glimpse of dynamic epistemic logic. Epistemological topics discussed include Moore-paradoxical phenomena, the surprise exam paradox, logical omniscience and epistemic closure, formalized theories of knowledge, debates about higher-order knowledge, and issues of knowability raised by Fitch's paradox. The references and recommended readings provide gateways for further exploration.

Keywords: epistemic modal logic, epistemology, epistemic closure, higher-order knowledge, knowability, epistemic paradoxes

1 Introduction

Once conceived as a single formal system, epistemic logic has become a general formal approach to the study of the structure of knowledge, its limits and possibilities, and its static and dynamic properties. In the 21st century there has been a resurgence of interest in the relation between epistemic logic and epistemology [Williamson, 2000, Sorensen, 2002, Hendricks, 2005, van Benthem, 2006, Stalnaker, 2006]. Some of the new applications of epistemic logic in epistemology go beyond the traditional limits of the logic of knowledge, either by modeling the dynamic process of knowledge acquisition or by modifying the representation of epistemic states to reflect different theories of knowledge. In this chapter, we begin with basic epistemic logic as it descends from Hintikka [1962] (§2-3), including multi-agent epistemic logic (§4) and doxastic logic (§5), followed by brief surveys of three topics at the interface of epistemic logic and epistemology: epistemic closure (§6), higher-order knowledge (§7), and knowability (§8).

2 Basic Models

Consider a simple formal language for describing the knowledge of an agent. The sentences of the language, which include all sentences of propositional logic, are generated from atomic sentences $p, q, r, \ldots$ using boolean connectives $\neg$ and $\wedge$

*Version of August 2015. The final version will appear as a chapter in the *Handbook of Formal Philosophy*, eds. S.O. Hansson and V.F. Hendricks, Springer, 2016.

(from which $\vee$, $\rightarrow$, and $\leftrightarrow$ are defined as usual) and a knowledge operator K.[1] We write that the agent knows that p as Kp, that she does *not* know that p *and* q as $\neg K(p \wedge q)$, that she knows *whether or not* q as $Kq \vee K\neg q$, that she knows that she does not know that *if* p*, then* q as $K\neg K(p \rightarrow q)$, and so on.

We interpret the language using a picture proposed by Hintikka [1962], which has since become familiar in philosophy. Lewis [1986] describes a version of the picture in terms of *ways the world might be*, compatible with one's knowledge:

> The content of someone's knowledge of the world is given by his class of *epistemically accessible* worlds. These are the worlds that might, for all he knows, be his world; world W is one of them iff he knows nothing, either explicitly of implicitly, to rule out the hypothesis that W is the world where he lives. (27)

The first part of the picture is that whatever is true in at least one of the agent's epistemically accessible worlds *might, for all the agent knows, be true in his world*, i.e., he does not know it to be false. The second part of the picture is that whatever is true in *all* of the agent's epistemically accessible worlds, the agent knows to be true, perhaps only implicitly (see Lewis 1986, §1.4).

Here we talk of "scenarios" rather than worlds, taking $w, v, u, \ldots$ to be scenarios and W to be a *set* of scenarios.[2] For our official definition of epistemic accessibility, call a scenario v epistemically accessible from a scenario w iff everything the agent knows in w is true in v [Williamson, 2000, §8.2].

Consider an example. A spymaster loses contact with one of his spies. In one of the spymaster's epistemically accessible scenarios, the spy has defected (d). In another such scenario, the spy remains loyal ($\neg d$). However, in all of the spymaster's epistemically accessible scenarios, the last message he received from the spy came a month ago (m). Hence the spymaster knows that the last message he received from the spy came a month ago, but he does not know whether or not the spy has defected, which we write as $Km \wedge \neg(Kd \vee K\neg d)$.

We assess the truth of such sentences in a *model* $\mathcal{M} = \langle W, R_K, V\rangle$, representing the epistemic state of an agent.[3] W is a nonempty set, the set of scenarios. R_K is a binary relation on W, such that for any w and v in W, we take wR_Kv to mean that scenario v is epistemically accessible from scenario w. Finally, V is a valuation function assigning to each atomic sentence p a subset of W, $V(p)$, which we take to be the set of scenarios in which p holds.

Given our definition of epistemic accessibility, and the fact that everything an agent *knows* is true, our intended models are ones in which R_K is *reflexive*: wR_Kw for all w in W. We call such models *epistemic models.*

Let φ and ψ be any sentences of the formal language. An atomic sentence p is *true* in a scenario w in a model $\mathcal{M} = \langle W, R_K, V\rangle$ iff w is in $V(p)$; $\neg\varphi$ is true

[1] To reduce clutter, I will not put quote marks around symbols and sentences of the formal language, trusting that no confusion will arise.

[2] In our formal models, "scenarios" will be unstructured points at which atomic sentences can be true or false. We are not committed to thinking of them as Lewisian possible worlds.

[3] Hintikka presented his original formal framework somewhat differently. Such details aside, we use the now standard relational structure semantics for normal modal logics.

in w iff φ is *not* true in w; $\varphi \wedge \psi$ is true in w iff φ and ψ are true in w; and finally, the modal clause matches both parts of the picture described above:

(MC) $K\varphi$ is true in w iff φ is true in every scenario v such that wR_Kv.

We say that a sentence is *satisfiable* iff it is true in some scenario in some model (otherwise *unsatisfiable*) and *valid* iff it is true in all scenarios in all models. We may also relativize these notions to a restricted class of models, such as the intended class of epistemic models in which R_K is reflexive. A sentence is *satisfiable in the class* iff it is true in some scenario in some model in the class and *valid over the class* iff it is true in all scenarios in all models in the class.

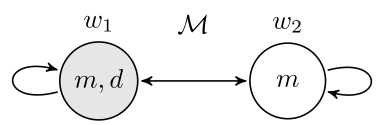

Figure 1: a simple epistemic model

Figure 1 displays a simple epistemic model for the spymaster example, where we draw a circle for each scenario (with all atomic sentences true in the scenario indicated inside the circle), and we draw an arrow from a scenario w to a scenario v iff wR_Kv. Observe that $Km \wedge \neg(Kd \vee K\neg d) \wedge d$ is true in w_1: d is true in w_1 by description; yet neither Kd nor $K\neg d$ is true in w_1, because neither d nor $\neg d$ is true in all scenarios epistemically accessible from w_1, namely w_2 and w_1 itself; however, Km is true in w_1, since m is true in all scenarios epistemically accessible from w_1. We could construct a more complicated epistemic model to represent the spymaster's knowledge and ignorance of other matters, but this simple model suffices to show that $Km \wedge \neg(Kd \vee K\neg d) \wedge d$ is satisfiable.

Let us now consider a sentence that is unsatisfiable in epistemic models. In a twist on Moore's [1942] paradox, Hintikka [1962, §4.17] considers what happens if I tell you something of the form *you don't know it, but the spy has defected*, translated as $d \wedge \neg Kd$. This may be true (as in w_1), but as Hintikka observes, you can never know it. You can never know that the spy has defected but you don't know it. Formally, $K(d \wedge \neg Kd)$ cannot be true in any scenario in an epistemic model; it is unsatisfiable, as we show in §3 below. It follows that $\neg K(d \wedge \neg Kd)$ is true in every scenario, so it is valid over epistemic models.

Since we take wR_Kv to mean that everything the agent knows in w is true in v, one might sense in (MC) some circularity or triviality. As a technical matter, there is no circularity, because R_K is a primitive in the model, not defined in terms of anything else. As a conceptual matter, we must be clear about the role of the epistemic model when paired with (MC): its role is to represent the content of one's knowledge, *what one knows*, not to analyze *what knowledge is* in terms of something else.[4] (As we discuss in §6 and §7, with richer epistemic

[4]It is important to draw a distinction between epistemic accessibility and other notions

structures we can also formalize such analyses of knowledge.) Finally, (MC) is not trivial because it is not neutral with respect to all theories of knowledge.[5]

3 Valid Principles

The reflexivity of R_K guarantees that the principle

$$\mathsf{T} \quad K\varphi \to \varphi$$

is valid.[6] For if $K\varphi$ is true in a scenario w, then by (MC), φ is true in all epistemically accessible scenarios, all v such that wR_Kv. Given wR_Kw by reflexivity, it follows that φ is true in w. (Conversely, if a relation R_K on a nonempty set W is not reflexive, then one can construct a model $\mathcal{M} = \langle W, R_K, V\rangle$ in which T is false. Thus, T *corresponds* to reflexivity.) It is also easy to verify that

$$\mathsf{M} \quad K(\varphi \wedge \psi) \to (K\varphi \wedge K\psi)$$

is valid over all models, simply by unpacking the truth definition. Using propositional logic (PL), we can now show why sentences of the Moorean form $p\wedge\neg Kp$ cannot be known:

(0) $K(p \wedge \neg Kp) \to (Kp \wedge K\neg Kp)$ instance of M;

of indistinguishability. Suppose that we replace R_K by a binary relation E on W, where our intuitive interpretation is that wEv holds "iff the subject's perceptual experience and memory" in scenario v "exactly match his perceptual experience and memory" in scenario w [Lewis, 1996, 553]. Suppose we were to then define the truth of $K\varphi$ in w as in (MC), but with R_K replaced by E. In other words, the agent knows φ in w iff φ is true in all scenarios that are experientially indistinguishable from w for the agent. (Of course, we could just as well reinterpret R_K in this way, without the new E notation.) There are two conceptual differences between the picture with E and the one with R_K. First, given the version of (MC) with E, the epistemic model with E does not simply represent the content of one's knowledge; rather, it commits us to a particular view of the conditions under which an agent has knowledge, specified in terms of perceptual experience and memory. Second, given our interpretation of E, it is plausible that E has certain properties, such as *symmetry* (wEv iff vEw), which are questionable as properties of R_K (see §7). Since the properties of the relation determine the valid principles for the knowledge operator K (as explained in §3 and §7), we must be clear about which interpretation of the relation we adopt: epistemic accessibility, experiential indistinguishability, or something else. Here we adopt the accessibility interpretation.

Finally, note that while one may read wR_Kv as "for all the agent knows in w, scenario v is the scenario he is in," one should *not* read wR_Kv as "in w, the agent considers scenario v possible," where the latter suggest a subjective psychological notion. The spymaster may not subjectively consider it possible that his spy, whom he has regarded for years as his most trusted agent, has defected. It obviously does not follow that he *knows* that his spy has not defected, as it would according to the subjective reading of R_K together with (MC).

[5] For any theory of knowledge that can be stated in terms of R_K and (MC), the rule RK of §3 must be sound. Therefore, theories for which RK is not sound, such as those discussed in §6, cannot be stated in this way. Given a formalization of such a theory, one can always define a relation R_K on scenarios such that wR_Kv holds iff everything the agent knows in w according to the formalization is true in v. It is immediate from this definition that if φ is not true in some v such that wR_Kv, then the agent does not know φ in w. However, it is *not* immediate that if φ is true in all v such that wR_Kv, then the agent knows φ in w. It is the right-to-left direction of (MC) that is not neutral with respect to all theories of knowledge.

[6] Throughout we use the nomenclature of modal logic for schemas and rules.

(1) $K\neg Kp \to \neg Kp$ instance of T;

(2) $K(p \wedge \neg Kp) \to (Kp \wedge \neg Kp)$ from (0)-(1) by PL;

(3) $\neg K(p \wedge \neg Kp)$ from (2) by PL.

The historical importance of this demonstration, now standard fare in epistemology, is that Hintikka explained a case of unknowability in terms of logical *form*. It also prepared the way for later formal investigations of Moorean phenomena (see van Ditmarsch et al. 2011 and refs. therein) in the framework of *dynamic epistemic logic*, discussed in §8.

To obtain a deductive system (**KT**) from which all and only the sentences valid over our reflexive epistemic models can be derived as theorems, it suffices to extend propositional logic with T and the following rule of inference:

$$\mathsf{RK} \quad \frac{(\varphi_1 \wedge \cdots \wedge \varphi_n) \to \psi}{(K\varphi_1 \wedge \cdots \wedge K\varphi_n) \to K\psi} \; (n \geq 0).$$

We interpret the rule to mean that if the sentence above the line is a theorem of the system, then the sentence below the line is also a theorem. Intuitively, RK says that the agent knows whatever follows logically from what she knows.

The soundness of RK shows that basic epistemic models involve a strong idealization. One can interpret these models as representing either the idealized (implicit, "virtual") knowledge of ordinary agents, or the ordinary knowledge of idealized agents (see Stalnaker 2006 and refs. therein). There is now a large literature on alternative models for representing the knowledge of agents with bounded rationality, who do not always "put two and two together" and therefore lack the *logical omniscience* reflected by RK (see Halpern and Pucella 2011 and refs. therein). As we discuss in §6 and §7, however, the idealized nature of our mathematical models can be beneficial in some philosophical applications.[7]

4 Multiple Agents

The formal language with which we began in §2 is the language of *single-agent* epistemic logic. The language of *multi-agent* epistemic logic contains an operator K_i for each agent i in a given set of agents. (We can also use these operators for different time-slices of the same agent, as shown below.) To interpret this language, we add to our models a relation R_{K_i} for each i, defining the truth of $K_i\varphi$ in a scenario w according to (MC) but with R_{K_i} substituted for R_K.

Suppose that the spymaster of §2, working for the KGB, is reasoning about the knowledge of a CIA spymaster. Consider two cases. In the first, although the KGB spymaster does not know whether his KGB spy has defected, he does know that the *CIA spymaster*, who currently has the upper hand, knows whether the KGB spy has defected. Model $\mathcal{N}$ in Figure 2 represents such a case, where the solid and dashed arrows are the epistemic accessibility relations for the KGB and CIA spymasters, respectively. The solid arrows for the KGB spymaster

[7]For additional ways of understanding idealization in epistemic logic, see Yap 2014.

between w_1 and w_2 indicate that his knowledge does not distinguish between these scenarios, whereas the absence of dashed arrows for the CIA spymaster between w_1 and w_2 indicates that her knowledge does distinguish between these scenarios, as the KGB spymaster knows. In the second case, by contrast, the KGB spymaster is uncertain not only about whether his KGB spy has defected, but also about whether the CIA spymaster knows whether the KGB spy has defected. Model $\mathcal{N}'$ in Figure 2 represents such a case. The KGB spymaster does not know whether he is in one of the upper scenarios, in which the CIA spymaster has no uncertainty, or one of the lower scenarios, in which the CIA spymaster is also uncertain about whether the KGB spy has defected. While $K_{\text{KGB}}(K_{\text{CIA}}d \vee K_{\text{CIA}}\neg d)$ is true in w_1 in $\mathcal{N}$, it is false in w_1 in $\mathcal{N}'$.

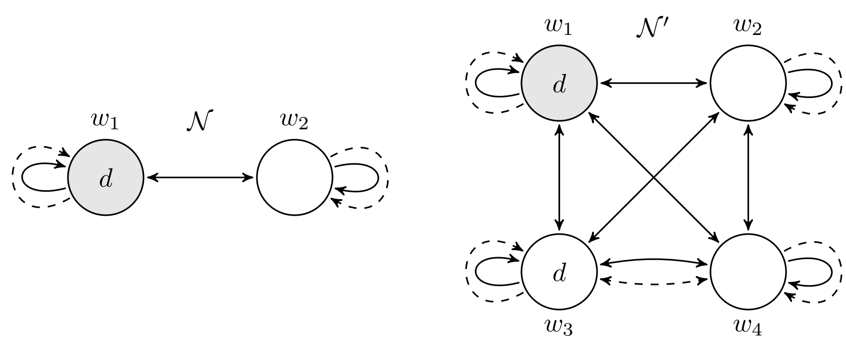

Figure 2: multi-agent epistemic models

Let us now turn from the representation of what agents know about the world and each other's knowledge, using multi-agent epistemic models, to formalized reasoning about such knowledge, using multi-agent epistemic logic.

For a sample application in epistemology, consider the *surprise exam paradox* (see Sorensen 1988 and refs. therein). A tutor announces to her student that she will give him a surprise exam at one of their daily tutoring sessions in the next n days, where an exam on day k is a surprise iff the student does not know on the morning of day k that there will be an exam that day. The student objects, "You can't wait until the last day, day n, to give the exam, because if you do, then I'll know on the morning of day n that the exam must be that day, so it won't be a surprise; since I can thereby eliminate day n, you also can't wait until day $n-1$ to give the exam, because if you do, then I'll know on the morning of day $n-1$ that the exam must be that day, so it won't be a surprise. . . ." Repeating this reasoning, he concludes that the supposed surprise exam cannot be on day $n-2$, day $n-3$, etc., or indeed on any day at all. His reasoning appears convincing. But then, as the story goes, the tutor springs an exam on him sometime before day n, and he is surprised. So what went wrong?

Consider the $n = 2$ case. For $i \in \{1, 2\}$, let e_i mean that the exam is on day i, and let $K_i\varphi$ mean that the student knows on the morning of day

i that φ, so our "multiple agents" are temporal stages of the student.[8] The tutor's announcement that there will be a surprise exam can be formalized as $(e_1 \wedge \neg K_1 e_1) \vee (e_2 \wedge \neg K_2 e_2)$. Now consider the following assumptions:

(A) $K_1((e_1 \wedge \neg K_1 e_1) \vee (e_2 \wedge \neg K_2 e_2))$;

(B) $K_1(e_2 \rightarrow K_2 \neg e_1)$;

(C) $K_1 K_2(e_1 \vee e_2)$.

Assumption (A) is that the student knows that the tutor's announcement of a surprise exam is true. Assumption (B) is that the student knows that he has a good memory: if the tutor waits until day 2 to give the exam, then the student will remember that it was not on day 1. Assumption (C) is that the student knows that he will also remember on the morning of day 2 that there was or will be an exam on one of the days (because, e.g., this is a school rule). The last assumption is that the student is a perfect logician in the sense of RK from §3. Let RK_i be the rule of inference just like RK but for the operator K_i. Then we can derive a Moorean absurdity from assumptions (A), (B), and (C):[9]

(4) $\big(K_2(e_1 \vee e_2) \wedge K_2 \neg e_1\big) \rightarrow K_2 e_2$ using PL and RK_2;

(5) $K_1\big(\big(K_2(e_1 \vee e_2) \wedge K_2 \neg e_1\big) \rightarrow K_2 e_2\big)$ from (4) by RK_1;

(6) $K_1(K_2 \neg e_1 \rightarrow K_2 e_2)$ from (C) and (5) using PL and RK_1;

(7) $K_1 \neg(e_2 \wedge \neg K_2 e_2)$ from (B) and (6) using PL and RK_1;

(8) $K_1(e_1 \wedge \neg K_1 e_1)$ from (A) and (7) using PL and RK_1.

We saw in §3 that sentences of the form of (8) are unsatisfiable in epistemic models, so we must give up either (A), (B), (C), or RK_i.[10] In this way, epistemic logic sharpens our options. We leave it to the reader to contemplate these options. There is much more to be said about the paradox (and the $n > 2$ case), but we have seen enough to motivate the interest of multi-agent epistemic logic.

The multi-agent setting also leads to the study of new epistemic concepts, such as *common knowledge* [Vanderschraaf and Sillari, 2013], but for the sake of space we return to the single-agent setting in the following sections.

[8] A similar formalization applies to the *designated student paradox* [Sorensen, 1988, 317], a genuinely multi-agent version of the surprise exam paradox.

[9] We skip steps for the sake of space. E.g., we obtain (4) by applying RK_2 to the tautology $((e_1 \vee e_2) \wedge \neg e_1) \rightarrow e_2$. We then obtain (5) directly from (4) using the special case of RK_1 where $n = 0$ in the premise $(\varphi_1 \wedge \cdots \wedge \varphi_n) \rightarrow \psi$, known as Necessitation: if ψ is a theorem, so is $K_1\psi$. It is important to remember that RK_i can only be applied to *theorems* of the logic, not to sentences that we have derived using undischarged assumptions like (A), (B), and (C). To be careful, we should keep track of the undischarged assumptions at each point in the derivation, but this is left to the reader as an exercise. Clearly we have not derived (8) as a theorem of the logic, since the assumptions (A), (B), and (C) are still undischarged. What we have derived as a theorem of the logic is the sentence abbreviated by $((A) \wedge (B) \wedge (C)) \rightarrow (8)$.

[10] We can derive (8) from (A), (B), and (C) in a doxastic logic (see §5) without the T axiom, substituting B_i for K_i. Thus, insofar as $B_1(e_1 \wedge \neg B_1 e_1)$ is also problematic for an ideal agent, the surprise exam paradox poses a problem about belief as well as knowledge.

5 Knowledge and Belief

The type of model introduced in §2 can represent not only the content of one's knowledge, but also the content of one's *beliefs*—and how these fit together. Let us extend the language of §2 with sentences of the form $B\varphi$ for belief and add to the models of §2 a *doxastic* accessibility relation R_B. We take wR_Bv to mean that everything the agent *believes* in w is true in v, and the truth clause for $B\varphi$ is simply (MC) with $K\varphi$ replaced by $B\varphi$ and R_K replaced by R_B. (For richer models representing *conditional* belief, see Stalnaker 1996, Board 2004.)

How do epistemic and doxastic accessibility differ? At the least, we should not require that R_B be reflexive, since it may not be that everything the agent believes in a scenario w is true in w. Instead, it is often assumed that R_B is *serial*: for all w, there is some v such that wR_Bv, some scenario where everything the agent believes is true. Given seriality, it is easy to see that the principle

$$\mathsf{D}\quad B\varphi \to \neg B\neg\varphi$$

is valid, in which case we are considering an agent with consistent beliefs. (Indeed, D *corresponds* to seriality in the same way that T corresponds to reflexivity, as noted in §3.) With or without seriality, the analogue of RK for belief,

$$\mathsf{RB}\quad \frac{(\varphi_1 \wedge \cdots \wedge \varphi_n) \to \psi}{(B\varphi_1 \wedge \cdots \wedge B\varphi_n) \to B\psi}\ (n \geq 0),$$

is also sound, an idealization that can be interpreted in ways analogous to those suggested for RK in §3, although RK raises additional questions (see §6).

How are epistemic and doxastic accessibility related? At the least, if whatever one knows one believes, then every scenario compatible with what one believes is compatible with what one knows: wR_Bv implies wR_Kv. Assuming this condition, $K\varphi \to B\varphi$ is valid; for if φ is true in all v such that wR_Kv, then by the condition, φ is true in all v such that wR_Bv. Other conditions relating R_B and R_K are often considered, reflecting assumptions about one's knowledge of one's beliefs and beliefs about one's knowledge (see Stalnaker 2006).

It is noteworthy in connection with Moore's [1942] paradox that if we make no further assumptions about the relation R_B, then $B(p \wedge \neg Bp)$ is *satisfiable*, in contrast to $K(p \wedge \neg Kp)$ from §3. In §7, we will discuss an assumption about R_B that is sometimes made and is sufficient to render $B(p \wedge \neg Bp)$ unsatisfiable.[11]

6 Epistemic Closure

The idealization that an agent knows whatever follows logically from what she knows raises two problems. In addition to the logical omniscience problem with RK noted in §3, there is a distinct objection to RK that comes from versions of the *relevant alternatives* (RA) [Dretske, 1970] and *truth-tracking* [Nozick, 1981] theories of knowledge. According to Dretske's [1970] theory, RK would fail

[11] In fact, the sentence $\neg B(p \wedge \neg Bp)$ precisely corresponds to a condition on R_B, namely that for every w, there is a v such that wR_Bv and for every u, vR_Bu implies wR_Bu.

even for "ideally astute logicians" who are "fully appraised of all the necessary consequences... of every proposition" (1010); even if RB were to hold for such an ideal logician, nonetheless RK would not hold for her in general. Nozick's [1981] theory leads to the same result. The reason is that one may satisfy the conditions for knowledge (ruling out the relevant alternatives, tracking the truth, etc.) with respect to some propositions and yet not with respect to all logical consequences of the set of those propositions, *even if* one has explicitly deduced all of the consequences. Hence the problem of epistemic closure raised by Dretske and Nozick is distinct from the problem of logical omniscience.

Dretske and Nozick famously welcomed the fact that their theories block appeals to the *closure of knowledge under known implication*,

K $(K\varphi \wedge K(\varphi \to \psi)) \to K\psi$,

in arguments for radical skepticism about knowledge.[12] For example, according to K, it is a necessary condition of an agent's knowing some mundane proposition p (Kp), e.g., that what she sees in the tree is a Goldfinch, that she knows that all sorts of skeptical hypotheses do not obtain ($K\neg\text{SH}$), e.g., that what she sees in the tree is not an animatronic robot, a hologram, etc., assuming she knows that these hypotheses are incompatible with p ($K(p \to \neg\text{SH})$). Yet it seems difficult or impossible to rule out every remote possibility raised by the skeptic. From here the skeptic reasons in reverse: since one has not ruled out every skeptical possibility, $K\neg\text{SH}$ is false, so given K and the truth of $K(p \to \neg\text{SH})$, it follows by PL that Kp is false. Hence we do not know mundane propositions about birds in trees—or almost anything else, as the argument clearly generalizes.

Rejecting the skeptical conclusion, Dretske and Nozick hold instead that K can fail. However, K is only one closure principle among (infinitely) many. Although Dretske [1970] denied K, he accepted other closure principles, such as closure under conjunction elimination, $K(\varphi \wedge \psi) \to (K\varphi \wedge K\psi)$, and disjunction introduction, $K\varphi \to K(\varphi \vee \psi)$. Nozick [1981] was prepared to give up even closure under conjunction elimination, but not closure under disjunction introduction. More generally, one can consider any closure principle of the form $(K\alpha_1 \wedge \cdots \wedge K\alpha_n) \to (K\beta_1 \vee \cdots \vee K\beta_m)$, such as $(Kp \wedge Kq) \to K(p \wedge q)$, $(K(p \vee q) \wedge K(p \to q)) \to Kq$, $K(p \wedge q) \to K(p \vee q)$, $K(p \wedge q) \to (Kp \vee Kq)$, etc.

To go beyond case-by-case assessments of closure principles, we can use an epistemic-logical approach to formalize theories of knowledge like those of Dretske, Nozick, and others, and then to obtain general characterizations of the valid closure principles for the formalized theories. To the extent that the formalizations are faithful, we can bring our results back to epistemology. For example, Holliday [2015] formalizes a family of RA and "subjunctivist" theories of knowledge using richer structures than the epistemic models in §2. The main Closure Theorem identifies exactly those closure principles of the form given above that are valid for the chosen RA and subjunctivist theories, with consequences for the closure debate in epistemology: on the one hand, the closure failures allowed by these theories spread far beyond those endorsed by Dretske

[12]Note that the K axiom is derivable from the RK rule with the tautology $(\varphi \wedge (\varphi \to \psi)) \to \psi$.

and Nozick; on the other hand, some closure principles that look about as useful to skeptics as K turn out to be valid according to these theories. While this result is negative for the theories in question, the formalization helps to identify the parameters of a theory of knowledge that affect its closure properties, clarifying the theory choices available to avoid the negative results.

As a methodological point, it is noteworthy that the results about epistemic closure in Holliday 2015, which tell us how RK fails for certain RA and subjunctivist theories of knowledge, apply to an agent whose beliefs satisfy full *doxastic closure* in the sense of RB. Thanks to this idealization, we can isolate failures of epistemic closure due to special conditions on knowledge, posited by a given epistemological theory, from failures of closure due to an agent's simply not "putting two and two together." This is an example of the beneficial role that idealization can play in epistemic logic, a point to which we return in §7.

7 Higher-Order Knowledge

Just as the T axiom $K\varphi \to \varphi$ corresponds to the reflexivity of R_K, other epistemic principles correspond to other conditions on R_K. In this way, our models give us another perspective on these principles via properties of accessibility.

First, consider *symmetry*: wR_Kv iff vR_Kw. Williamson [2000, §8.2] observes that this assumption plays a crucial role in some arguments for radical skepticism about knowledge. Suppose that in scenario w, the agent has various true beliefs about the external world. The skeptic describes a scenario v in which those beliefs are false, but the agent is systematically deceived into holding them anyway. How does one know that one is not in such a scenario? Uncontroversially, it is compatible with everything the agent knows in the skeptical scenario v that she is in the ordinary scenario w. Given this, the skeptic appeals to symmetry: it must then be compatible with everything the agent knows in w that she is in v, which is to say that everything she knows in w is true in v. But since everything the agent believes in w about the external world is *false* in v, the skeptic concludes that such beliefs do not constitute knowledge in w.

If we require with the skeptic that R_K be symmetric, then the principle

B $\neg\varphi \to K\neg K\varphi$

is valid according to (MC).[13] Although this is often assumed for convenience in applications of epistemic logic in computer science and game theory, the validity of B is clearly too strong as a matter of epistemology (see Williamson 2000).[14]

It is easy to check that symmetry follows if R_K is both reflexive and *Euclidean*: if wR_Kv and wR_Ku, then vR_Ku. The latter property guarantees that

5 $\neg K\varphi \to K\neg K\varphi$

[13] Assume $\neg\varphi$ is true in w, so φ is not true in w. Consider some v with wR_Kv. By symmetry, vR_Kw. Then since φ is not true in w, $K\varphi$ is not true in v by (MC), so $\neg K\varphi$ is true in v. Since v was arbitrary, $\neg K\varphi$ is true in all v such that wR_Kv, so $K\neg K\varphi$ is true in w by (MC).

[14] Note that if we reject the requirement that R_K be symmetric in *every* epistemic model, we can still allow models in which R_K is symmetric (such as the model in Figure 1), when this is appropriate to model an agent's knowledge. The same applies for other properties.

is valid according to (MC). Hence if we reject the symmetry requirement and the validity of the B axiom, which corresponds to symmetry, then we must also reject the Euclidean requirement and the validity of the 5 axiom, which corresponds to Euclideanness. Additional arguments against the 5 axiom come from considering the interaction of knowledge and belief (recall §5).[15]

While the rejection of B and 5 is universal among epistemologists, there is another principle of higher-order knowledge defended by some. Corresponding to the condition that R_K is *transitive* (if wR_Kv and vR_Ku, then wR_Ku) is the principle

4 $K\varphi \to KK\varphi$.[16]

Similarly, corresponding to the condition that R_B is transitive is the principle $B\varphi \to BB\varphi$. Assuming the latter, $B(p \wedge \neg Bp)$ is unsatisfiable, which is the fact at the heart of Hintikka's [1962, §4.6-4.7] analysis of Moore's paradox.[17]

Hintikka [1962, §5.3] argued that 4 holds for a strong notion of knowledge, found in philosophy from Aristotle to Schopenhauer. The principle has since become known in epistemology as "KK" and in epistemic logic as "positive introspection." Yet Hintikka [1962, §3.8-3.9, §5.3-5.4] rejected arguments for 4 based on claims about agents' introspective powers, or what he called "the myth of the self-illumination of certain mental activities" (67). Instead, his claim was that for a strong notion of knowledge, *knowing that one knows* "differs only in words" from *knowing*. His arguments for this claim [1962, §2.1-2.2] deserve further attention, but we cannot go into them here (see Stalnaker 1996, §1).

As Hintikka assumed only reflexivity and transitivity for R_K, his investigation of epistemic logic settled on the modal logic of reflexive and transitive models, **S4**, obtained by extending propositional logic with RK, T, and 4. Some objected to this proposal on the grounds that given $K\varphi \to B\varphi$, 4 implies $K\varphi \to BK\varphi$, which invites various counterexamples (see the articles in *Synthese*, Vol. 21, No. 2, 1970). Rejecting these objections, Lenzen [1978, Ch. 4] argued from considerations of the combined logic of knowledge and belief (and "conviction") that the logic of knowledge is at least as strong as a system extending **S4** known as **S4.2** and at most as strong as one known as **S4.4**. Others implicated 4 in the surprise exam paradox, while still others argued for 4's innocence (see Williamson 2000, Ch. 6 and Sorensen 1988, Ch. 7-8).

[15] Assuming $K\varphi \to B\varphi$, D, and 5, the principle $BK\varphi \to K\varphi$ is derivable (see Gochet and Gribomont 2006, §2.4). Given the same assumptions, if an agent is a "stickler" [Nozick, 1981, 246] who believes something only if she believes that she knows it ($B\varphi \to BK\varphi$), then one can even derive $B\varphi \leftrightarrow K\varphi$ (see Lenzen 1978 and Halpern 1996). Given $K\varphi \to B\varphi$, D, B, and $B\varphi \to BK\varphi$, one can still derive $B\varphi \to \varphi$ (see Halpern 1996, 485).

[16] To see that 4 is valid over the class of transitive models, assume that $K\varphi$ is true in w in such a model, so by (MC), φ is true in all v such that wR_Kv. Consider some u with wR_Ku. Toward proving that $K\varphi$ is true in u, consider some v with uR_Kv. By transitivity, wR_Ku and uR_Kv implies wR_Kv. Hence by our initial assumption, φ is true in v. Since v was arbitrary, φ is true in all v such that uR_Kv, so $K\varphi$ is true in u by (MC). Finally, since u was arbitrary, $K\varphi$ is true in all u such that wR_Ku, so $KK\varphi$ is true in w by (MC).

[17] Assuming D, 4, and M for B, we have: (i) $B(p \wedge \neg Bp)$, assumption for reductio; (ii) $Bp \wedge B\neg Bp$, from (i) by M for B and PL; (iii) $BBp \wedge B\neg Bp$, from (ii) by 4 for B and PL; (iv) $\neg B\neg Bp \wedge B\neg Bp$, from (iii) by D and PL; (v) $\neg B(p \wedge \neg Bp)$, from (i)-(iv) by PL.

In addition to approaching questions of higher-order knowledge via properties of R_K, we can approach these questions by formalizing substantive theories of knowledge. While the relevant alternatives and subjunctivist theories mentioned in §6 are generally hostile to 4, other theories are friendlier to 4. For example, consider what Stalnaker [1996] calls the *defeasibility analysis*: "define knowledge as belief (or justified belief) that is stable under any potential revision by a piece of information that is in fact true" (187). Like others, Stalnaker [2006] finds such stability too strong as a necessary condition for knowledge; yet he finds its sufficiency more plausible. (Varieties of belief stability have since been studied for their independent interest, e.g., in Baltag and Smets 2008, without commitment to an analysis of knowledge.) Formalizing the idea of stability under belief revision in models encoding agents' *conditional* beliefs, Stalnaker [1996, 2006] shows that under some assumptions about agents' access to their own conditional beliefs, the formalized defeasibility analysis validates 4.[18]

The most influential recent contribution to the debate over 4 is Williamson's [1999, 2000, Ch. 5] *margin of error* argument, which we will briefly sketch. Consider a perfectly rational agent who satisfies the logical omniscience idealization of RK and hence K, setting aside for now the additional worries about closure raised in §6. Williamson argues that even for such an agent, 4 does not hold in general. Suppose the agent is estimating the height of a faraway tree, which is in fact k inches. Let h_i stand for *the height of the tree is i inches*, so h_k is true. While the agent's rationality is perfect, his eyesight is not. As Williamson [2000, 115] explains, "anyone who can tell by looking at the tree that it is not i inches tall, when in fact it is $i+1$ inches tall, has much better eyesight and a much greater ability to judge heights" than this agent. Hence for any i, we have $h_{i+1} \to \neg K \neg h_i$. In contrapositive form, this is equivalent to:

(9) $\forall i (K\neg h_i \to \neg h_{i+1})$.[19]

Now suppose that the agent reflects on the limitations of his visual discrimination and comes to know every instance of (9), so that the following holds:

(10) $\forall i (K(K\neg h_i \to \neg h_{i+1}))$.

Given these assumptions, it follows that for any j, if the agent knows that the height is not j inches, then he also knows that the height is not $j+1$ inches:

(11) $K\neg h_j$ assumption;

(12) $KK\neg h_j$ from (11) using 4 and PL;

(13) $K(K\neg h_j \to \neg h_{j+1})$ instance of (10);

(14) $K\neg h_{j+1}$ from (12) and (13) using K and PL.

[18] Stalnaker shows that the epistemic logic of the defeasibility analysis as formalized is **S4.3**, which is intermediate in strength between Lenzen's lower and upper bounds of **S4.2** and **S4.4**.

[19] Note that the universal quantifiers in (9), (10), (15), and (21) are not part of our formal language. They are merely shorthand to indicate a schema of sentences.

Assuming the agent knows that the tree's height is not 0 inches, so $K\neg h_0$ holds, by repeating the steps of (11)-(14), we reach the conclusion $K\neg h_k$ by induction. (We assume, of course, that the agent has the appropriate beliefs implied by (11)-(14), as a result of following out the consequences of what he knows.) Finally, by T, $K\neg h_k$ implies $\neg h_k$, contradicting our initial assumption of h_k.

Williamson concludes that this derivation of a contradiction is a reductio ad absurdum of $\mathsf{4}$. Rejecting the transitivity of epistemic accessibility, he proposes formal models of knowledge with non-transitive accessibility to model limited discrimination [Williamson, 1999]. (For discussion, see *Philosophy and Phenomenological Research*, Vol. 64, No. 1, 2002, and a number of recent papers by Bonnay and Egré, e.g., Bonnay and Egré 2009. Williamson [2014] goes further and argues that an agent can know a proposition p even though the probability on her evidence that she knows p is as close to 0 as we like.) Since Williamson's argument assumes that the agent satisfies the idealization given by RK in §2, if it is indeed a reductio of $\mathsf{4}$ in particular, then it shows that $\mathsf{4}$ fails for reasons other than bounded rationality. As Williamson suggests (see Hendricks and Roy 2010, Ch. 25), this shows how idealization in epistemic logic can play a role analogous to that of idealization in science, allowing one to better discern the specific effects of a particular phenomenon such as limited discrimination.

8 Knowability

We now turn from questions about epistemic closure and higher-order knowledge to questions about the limits of what one may come to know. As we will see, these questions lead naturally to a *dynamic* approach to epistemic logic.

Fitch [1963] derived an unexpected consequence from the thesis, advocated by some anti-realists, that *every truth is knowable*. Let us express this thesis as

(15) $\forall q(q \to \Diamond Kq)$,

where $\Diamond$ is a *possibility* operator. Fitch's proof uses the two modest assumptions about K used for (0)-(3) in §3, T and M, together with two modest assumptions about $\Diamond$. First, $\Diamond$ is the dual of a *necessity* operator $\Box$ such that $\neg\Diamond\varphi$ follows from $\Box\neg\varphi$. Second, $\Box$ obeys the rule of Necessitation: if φ is a theorem, then $\Box\varphi$ is a theorem. For an arbitrary p, consider the following:

(16) $(p \wedge \neg Kp) \to \Diamond K(p \wedge \neg Kp)$ instance of (15).

Since we demonstrated in §3 that $\neg K(p \wedge \neg Kp)$ is a theorem, we have:

(17) $\Box\neg K(p \wedge \neg Kp)$ from (0)-(3) by Necessitation;

(18) $\neg\Diamond K(p \wedge \neg Kp)$ from (17) by duality of $\Diamond$ and $\Box$;

(19) $\neg(p \wedge \neg Kp)$ from (16) and (18) by PL;

(20) $p \to Kp$ from (19) by (classical) PL;

(21) $\forall p(p \to Kp)$ from (16)-(20), since p was arbitrary.

From the original anti-realist assumption in (15) that every truth is *knowable*, it follows in (21) that every truth is *known*, an absurd conclusion.

There is now a large literature devoted to this "knowability paradox" (see, e.g., Williamson 2000, Ch. 12, Edgington 1985, Sorensen 1988, Ch. 4, and Salerno 2009). There are proposals for blocking the derivation of (21) at various places, e.g., in the step from (19) to (20), which is not valid in *intuitionistic* logic, or in the universal instantiation step in (16), since it allegedly involves an illegitimate substitution into an intensional context. Yet another question raised by Fitch's proof concerns how we should interpret the $\Diamond$ operator in (15).

Van Benthem [2004] proposes an interpretation of the $\Diamond$ in the framework of dynamic epistemic logic (see van Benthem 2011 and the chapter of this Handbook by Baltag and Smets for refs.). As we state more formally below, the idea is that $\Diamond K\varphi$ is true iff *there is a possible change in one's epistemic state* after which one knows φ. Contrast this with the *metaphysical* interpretation of $\Diamond$, according to which $\Diamond K\varphi$ is true iff there is a possible world where one knows φ.

In the simplest dynamic approach, we model a change in an agent's epistemic state as an *elimination of epistemic possibilities.* Recall the spymaster example from §2. We start with an epistemic model $\mathcal{M}$ and an actual scenario w_1, representing the spymaster's initial epistemic state. Although his spy has defected, initially the spymaster does not know this, so $d \wedge \neg Kd$ is true in w_1 in $\mathcal{M}$. Suppose the spymaster then learns the news of his spy's defection. To model this change in his epistemic state, we *eliminate* from $\mathcal{M}$ all scenarios in which d is *not* true, resulting in a *new* epistemic model $\mathcal{M}_{|d}$, displayed in Figure 3, which represents the spymaster's new epistemic state. Note that Kd is true in w_1 in $\mathcal{M}_{|d}$, reflecting the spymaster's new knowledge of his spy's defection.

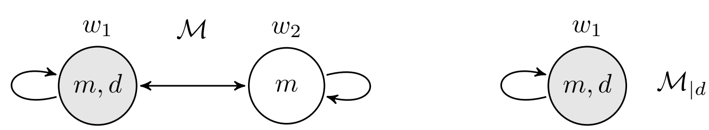

Figure 3: modeling knowledge acquisition by elimination of possibilities

The acquisition of knowledge is not always as straightforward as just described. Suppose that instead of learning d, the spymaster is informed that *you don't know it, but the spy has defected*, the familiar $d \wedge \neg Kd$. The resulting model $\mathcal{M}_{|d\wedge\neg Kd}$, obtained by eliminating from $\mathcal{M}$ all scenarios in which $d\wedge\neg Kd$ is false (namely w_2) is the same as $\mathcal{M}_{|d}$ in this case. However, while $d \wedge \neg Kd$ is true in w_1 in $\mathcal{M}$, it becomes *false* in w_1 in $\mathcal{M}_{|d\wedge\neg Kd}$, since Kd becomes true in w_1 in $\mathcal{M}_{|d\wedge\neg Kd}$. As Hintikka [1962] observes of a sentence like $d \wedge \neg Kd$, "If you know that I am well informed and if I address the words ... to you," then you "may come to know that what I say *was* true, but saying it in so many words has the effect of making what is being said false" (68f).[20] Since $d \wedge \neg Kd$ is false

[20]For discussion of such "unsuccessful" announcements in the context of the surprise exam paradox, see Gerbrandy 2007.

in w_1 in $\mathcal{M}_{|d\wedge\neg Kd}$, so is $K(d \wedge \neg Kd)$.

Returning to the knowability paradox, van Benthem's proposal, stated informally above, is to interpret the $\Diamond$ in (15) such that $\Diamond K\varphi$ is true in a scenario w in a model $\mathcal{M}$ iff *there exists* some ψ true in w such that $K\varphi$ is true in w in the model $\mathcal{M}_{|\psi}$, obtained by eliminating from $\mathcal{M}$ all scenarios in which ψ is false. For example, in Figure 3, $\Diamond Kd$ is true in w_1 in $\mathcal{M}$, since we may take d itself for the sentence ψ; but $\Diamond K(d \wedge \neg Kd)$ is false, since there is no ψ that will get the spymaster to know $d \wedge \neg Kd$. As expected, (15) is not valid for all sentences on this interpretation of $\Diamond$. Yet we now have a formal framework (see Balbiani et al. 2008) in which to investigate the sentences for which (15) *is* valid.

A much-discussed proposal by Tennant (see Salerno 2009, Ch. 14) is to restrict (15) to apply only to *Cartesian* sentences, those φ such that $K\varphi$ is consistent, in the sense that one cannot derive a contradiction from $K\varphi$. This restriction blocks the substitution of $p \wedge \neg Kp$, given (0)-(3) in §3. However, van Benthem [2004] shows that (15) is not valid for all Cartesian sentences on the dynamic interpretation of $\Diamond$, which imposes stricter constraints on knowability. Another conjecture is that the sentences for which (15) is valid on the dynamic interpretation of $\Diamond$ are those that one can always learn without *self-refutation*, in the sense of Hintikka's remark above. Surprisingly, this conjecture is false, as there are sentences φ such that whenever φ is true, one can come to know φ by being informed of *some* true ψ, but one cannot always come to know φ by being informed of φ *itself* [van Benthem, 2004]. A syntactic characterization of the sentences for which (15) is valid on the dynamic interpretation of $\Diamond$ is currently unknown, an open problem for future research (see van Ditmarsch et al. 2011 for another sense of "everything is knowable"). We conclude by observing that while Fitch's proof may make trouble for anti-realism, reframing the issue in terms of the dynamics of knowledge acquisition opens a study of positive lessons about knowability (see van Benthem 2004, §8; cf. Williamson 2000, §12.1).

9 Conclusion

This survey has given only a glimpse of the intersection of epistemic logic and epistemology. Beyond its scope were applications of epistemic logic to epistemic paradoxes besides the surprise exam (see Sorensen 2014), to debates about fallibilism and contextualism in epistemology (see refs. in Holliday 2015), to Gettier cases [Williamson, 2013], and to social epistemology. Also beyond the scope of this survey were systems beyond basic epistemic logic, including quantified epistemic logic,[21] justification logic [Artemov, 2008], modal operator epistemology [Hendricks, 2005], and logics of group knowledge [Vanderschraaf and Sillari, 2013]. For a sense of where leading figures in the intersection foresee progress, we refer the reader to Hendricks and Roy 2010 and Hendricks and Pritchard

[21]See Hintikka 1962, Ch. 6, Gochet and Gribomont 2006, §5, and Aloni 2005 for discussion of quantified epistemic logic. Hintikka [2003] has proposed a "second generation" epistemic logic, based on *independence-friendly* first-order logic, aimed at solving the difficulties and fulfilling the epistemological promises of quantified epistemic logic.

2008. Given the versatility of contemporary epistemic logic, the prospects for fruitful interactions with epistemology are stronger than ever before.[22]

References and Recommended Readings

Maria Aloni. Individual Concepts in Modal Predicate Logic. *Journal of Philosophical Logic*, 34(1):1–64, 2005.
*** For a modern treatment of classic puzzles of quantified epistemic logic, go here next.

Sergei Artemov. The Logic of Justification. *The Review of Symbolic Logic*, 1 (4):477–513, 2008.

Philippe Balbiani, Alexandru Baltag, Hans van Ditmarsch, Andreas Herzig, Tomohiro Hoshi, and Tiago de Lima. 'Knowable' as 'Known after an Announcement'. *The Review of Symbolic Logic*, 1:305–334, 2008.

Alexandru Baltag and Sonja Smets. A Qualitative Theory of Dynamic Belief Revision. In Giacomo Bonanno, Wiebe van der Hoek, and Michael Wooldridge, editors, *Logic and the Foundations of Game and Decision Theory*, Texts in Logic and Games 3, pages 13–60. Amsterdam University Press, Amsterdam, 2008.

Johan van Benthem. What One May Come to Know. *Analysis*, 64(2):95–105, 2004.
*** For further philosophical introduction to dynamic epistemic logic, go here next. A longer version appears as Chapter 9 of Salerno 2009.

Johan van Benthem. Epistemic Logic and Epistemology: The State of Their Affairs. *Philosophical Studies*, 128(1):49–76, 2006.

Johan van Benthem. *Logical Dynamics of Information and Interaction.* Cambridge University Press, New York, 2011.

Oliver Board. Dynamic Interactive Epistemology. *Games and Economic Behavior*, 49:49–80, 2004.

Denis Bonnay and Paul Egré. Inexact Knowledge with Introspection. *Journal of Philosophical Logic*, 38:179–227, 2009.

Hans van Ditmarsch, Wiebe van der Hoek, and Peter Iliev. Everything is Knowable – How to Get to Know *Whether* a Proposition is True. *Theoria*, 78(2): 93–114, 2011.

Fred Dretske. Epistemic Operators. *The Journal of Philosophy*, 67(24):1007–1023, 1970.
*** A classic source of current debates about epistemic closure.

[22]For helpful discussion and comments, I wish to thank Johan van Benthem, Tomohiro Hoshi, Thomas Icard, Alex Kocurek, Eric Pacuit, John Perry, Igor Sedlár, Justin Vlasits, and the students in my Fall 2012 seminar on Epistemic Logic and Epistemology at UC Berkeley.

Dorothy Edgington. The Paradox of Knowability. *Mind*, 94(376):557–568, 1985.

Paul Ègre. Epistemic Logic. In Leon Horsten and Richard Pettigrew, editors, *The Continuum Companion to Philosophical Logic*, Continuum Companions, pages 503–542. Continuum, New York, 2011.
*** For an introduction to further topics in epistemic logic, go here next.

Frederic B. Fitch. A Logical Analysis of Some Value Concepts. *The Journal of Symbolic Logic*, 28(2):135–142, 1963.

Jelle Gerbrandy. The Surprise Examination in Dynamic Epistemic Logic. *Synthese*, 155:21–33, 2007.

Paul Gochet and Pascal Gribomont. Epistemic Logic. In Dov M. Gabbay and John Woods, editors, *Handbook of the History of Logic*, volume 7. Elsevier, Amsterdam, 2006.

Joseph Y. Halpern. Should Knowledge Entail Belief? *Journal of Philosophical Logic*, 25:483–494, 1996.

Joseph Y. Halpern and Riccardo Pucella. Dealing with Logical Omniscience: Expressiveness and Pragmatics. *Artificial Intelligence*, 175:220–235, 2011.

Vincent F. Hendricks. *Mainstream and Formal Epistemology.* Cambridge University Press, New York, 2005.

Vincent F. Hendricks and Duncan Pritchard, editors. *Epistemology: 5 Questions.* Automatic Press, Copenhagen, 2008.

Vincent F. Hendricks and Olivier Roy, editors. *Epistemic Logic: 5 Questions.* Automatic Press, Copenhagen, 2010.

Jaakko Hintikka. *Knowledge and Belief: An Introduction to the Logic of the Two Notions.* Cornell University Press, Ithaca, 1962.
*** The founding text of modern epistemic logic, still worth reading.

Jaakko Hintikka. A Second Generation Epistemic Logic and its General Significance. In Vincent F. Hendricks, Klaus Frovin Jørgensen, and Stig Andur Pedersen, editors, *Knowledge Contributors*, volume 322 of *Synthese Library.* Kluwer, Dordrecht, 2003.

Wesley H. Holliday. Epistemic Closure and Epistemic Logic I: Relevant Alternatives and Subjunctivism. *Journal of Philosophical Logic*, 44(1):1–62, 2015.

Wolfgang Lenzen. Recent Work in Epistemic Logic. *Acta Philosophica Fennica*, 30:1–219, 1978.

David Lewis. *On the Plurality of Worlds.* Basil Blackwell, Oxford, 1986.

David Lewis. Elusive Knowledge. *Australasian Journal of Philosophy*, 74(4): 549–567, 1996.

G. E. Moore. A Reply to My Critics. In Paul A. Schilpp, editor, *The Philosophy of G.E. Moore*, pages 535–677. Northwestern University Press, Evanston, 1942.

Robert Nozick. *Philosophical Explanations*. Harvard University Press, Cambridge, Mass., 1981.

Eric Pacuit. Dynamic Epistemic Logic I and II. *Philosophy Compass*, 8(9): 798–833, 2013.

Joe Salerno, editor. *New Essays on the Knowability Paradox*. Oxford University Press, New York, 2009.

Roy Sorensen. *Blindspots*. Clarendon Press, Oxford, 1988.
*** Filled with fascinating material relevant to epistemic logic.

Roy Sorensen. Formal Problems About Knowledge. In Paul K. Moser, editor, *The Oxford Handbook of Epistemology*, pages 539–595. Oxford University Press, New York, 2002.

Roy Sorensen. Epistemic Paradoxes. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Spring 2014 edition, 2014.

Robert Stalnaker. Knowledge, Belief and Counterfactual Reasoning in Games. *Economics and Philosophy*, 12:133–162, 1996.

Robert Stalnaker. On Logics of Knowledge and Belief. *Philosophical Studies*, 128(1):169–199, 2006.
*** For more on the interaction of knowledge and belief, go here next.

Gail C. Stine. Skepticism, Relevant Alternatives, and Deductive Closure. *Philosophical Studies*, 29(4):249–261, 1976.

Peter Vanderschraaf and Giacomo Sillari. Common Knowledge. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Fall 2013 edition, 2013.

Timothy Williamson. Rational Failures of the KK Principle. In Cristina Bicchieri, Richard C. Jeffrey, and Brian Skyrms, editors, *The Logic of Strategy*, pages 101–118. Oxford University Press, New York, 1999.
*** For payoffs of applying epistemic logic to epistemology, start here or with one of Williamson's other papers.

Timothy Williamson. *Knowledge and its Limits*. Oxford University Press, New York, 2000.

Timothy Williamson. Gettier Cases in Epistemic Logic. *Inquiry*, 56(1):14, 2013.

Timothy Williamson. Very Improbable Knowing. *Erkenntnis*, 79(5):971–999, 2014.

Audrey Yap. Idealization, Epistemic Logic, and Epistemology. *Synthese*, 191 (14):3351–3366, 2014.