J Philos Logic (2015) 44:1–62
DOI 10.1007/s10992-013-9306-2

Epistemic Closure and Epistemic Logic I: Relevant Alternatives and Subjunctivism

Wesley H. Holliday

Received: 10 April 2012 / Accepted: 23 September 2013 / Published online: 5 April 2014

Abstract Epistemic closure has been a central issue in epistemology over the last forty years. According to versions of the *relevant alternatives* and *subjunctivist* theories of knowledge, epistemic closure can fail: an agent who knows some propositions can fail to know a logical consequence of those propositions, even if the agent explicitly believes the consequence (having "competently deduced" it from the known propositions). In this sense, the claim that epistemic closure can fail must be distinguished from the fact that agents do not always believe, let alone know, the consequences of what they know—a fact that raises the "problem of logical omniscience" that has been central in epistemic logic. This paper, part I of II, is a study of epistemic closure from the perspective of epistemic logic. First, I introduce models for epistemic logic, based on Lewis's models for counterfactuals, that correspond closely to the pictures of the relevant alternatives and subjunctivist theories of knowledge in epistemology. Second, I give an exact characterization of the closure properties of knowledge according to these theories, as formalized. Finally, I consider the relation between closure and higher-order knowledge. The philosophical repercussions of these results and results from part II, which prompt a reassessment of the issue of closure in epistemology, are discussed further in companion papers. As a contribution to modal logic, this paper demonstrates an alternative approach to proving modal completeness theorems, without the standard canonical model construction. By "modal decomposition" I obtain completeness and other results for two non-normal modal logics with respect to new semantics. One of these logics, dubbed *the logic of ranked relevant alternatives*, appears not to have been previously identified in the modal logic literature. More broadly, the paper presents epistemology as a rich area for logical study.

W. H. Holliday (✉)
Department of Philosophy, University of California, 314 Moses Hall #2390,
Berkeley, CA 94720-2390, USA
e-mail: wesholliday@berkeley.edu

Keywords Epistemic closure · Epistemic logic · Epistemology · Modal logic · Relevant alternatives · Safety · Sensitivity · Subjunctivism · Tracking

1 Introduction

The debate over epistemic closure has been called "one of the most significant disputes in epistemology over the last forty years" [45, 256]. The starting point of the debate is typically some version of the claim that knowledge is *closed under known implication* (see Dretske [22]). At its simplest, it is the claim that if an agent knows φ and knows that φ implies ψ, then the agent knows ψ: $(K\varphi \wedge K(\varphi \to \psi)) \to K\psi$, in the language of epistemic logic.

An obvious objection to the simple version of the claim is that an agent with bounded rationality may know φ and know that φ implies ψ, yet not "put two and two together" and draw a conclusion about ψ. Such an agent may not even believe ψ, let alone know it. The challenge of the much-discussed "problem of logical omniscience" (see, e.g., Stalnaker [69]; Halpern and Pucella [29]) is to develop a good theoretical model of the knowledge of such agents.

According to a different objection, made famous in epistemology by Dretske [19] and Nozick [58] (and applicable to more sophisticated closure claims), knowledge would not be closed under known implication even for "ideally astute logicians" [19, 1010] who always put two and two together and believe all consequences of what they believe. This objection (explained in Section 2), rather than the logical omniscience problem, will be our starting point.[1]

The closure of knowledge under known implication, henceforth referred to as 'K' after the modal axiom given above, is one closure principle among infinitely many. Although Dretske [19] denied K, he accepted other closure principles, such as closure under conjunction elimination, $K(\varphi\wedge\psi) \to K\varphi$, and closure under disjunction introduction, $K\varphi \to K(\varphi \vee \psi)$ (1009). By contrast, Nozick [58] was prepared to give up closure under conjunction elimination (228), although not closure under disjunction introduction (230n64, 692).

Dretske and Nozick not only provided examples in which they claimed K fails, but also proposed theories of knowledge that they claimed would explain the failures, as discussed below. Given such a theory, one may ask: is the theory committed to the failure of other, weaker closure principles, such as those mentioned above? Is it committed to closure failures in situations other than those originally envisioned as counterexamples to K? The concern is that closure failures may spread, and they may spread to where no one wants them.

Pressing such a *problem of containment* has an advantage over other approaches to the debate over K. It appeals to considerations that both sides of the debate are likely

[1] Other epistemologists who have denied closure under known implication in the relevant sense include McGinn [55], Goldman [27], Audi [4], Heller [34], Harman and Sherman [31, 65], Lawlor [47], Becker [7], and Adams et al. [1].

to accept, rather than merely insisting on the plausibility of K (or of one of its more sophisticated versions). A clear illustration of this approach is Kripke's [44] barrage of arguments to the effect that closure failures are ubiquitous given Nozick's theory of knowledge. In a different way, Hawthorne [32, 41] presses the first part of the containment problem against Dretske and Nozick, as I critically discuss in Holliday [38, Section 6.1.2].[2]

In this paper, I formally assess the problem of containment for a family of prominent "modal" theories of knowledge (see, e.g., Pritchard [61]; Black [9]). In particular, I introduce formal models of the following: the *relevant alternatives* (RA) theories of Lewis [52] and Heller [33, 34]; one way of developing the RA theory of Dretske [21] (based on Heller); the basic *tracking* theory of Nozick [58]; and the basic *safety* theory of Sosa [67]. A common feature of the theories of Heller, Nozick, and Sosa, which they share with those of Dretske [20], Goldman [26], and others, is some subjunctive or counterfactual-like condition(s) on knowledge, relating what an agent knows to what holds in selected *counterfactual possibilities* or *epistemic alternatives*.

Vogel [76] characterizes *subjunctivism* as "the doctrine that what is distinctive about knowledge is essentially modal in character, and thus is captured by certain subjunctive conditionals" (73), and some versions of the RA theory have a similar flavor.[3] I will call this family of theories *subjunctivist flavored*. Reflecting their commonality, my formal framework is based on the formal semantics for subjunctive conditionals in the style of Lewis [49] and Stalnaker [68]. As a result, the epistemic logics studied here behave very differently than traditional epistemic logics in the style of Hintikka [36]. (For a philosophically-oriented review of basic epistemic logic, see Holliday [39]).

This paper is part I of II. The main result of part I is an exact characterization in propositional epistemic logic of the closure properties of knowledge according to the RA, tracking, and safety theories, as formalized. Below I preview some of the epistemological and logical highlights of this and other results from part I. Part II introduces a unifying framework in which all of the theories of knowledge studied here fit as special cases; I argue that the closure problems with these theories are symptoms of inherent problems in their framework; and I propose to solve these problems with a new framework for *fallibilist* theories of knowledge. Elsewhere I discuss the philosophical repercussions of the results from parts I and II in depth [38, 40].

[2] Lawlor [47, 44] makes the methodological point about the advantage of raising the containment problem. It is noteworthy that Hawthorne takes a kind of proof-theoretic approach; he argues that a certain set of closure principles, not including K, suffices to derive the consequences that those who deny K wish to avoid, in which case they must give up a principle in the set. By contrast, our approach will be model-theoretic; we will study models of particular theories to identify those structural features that lead to closure failures.

[3] The view that knowledge has a modal character and the view that it is captured by subjunctive conditionals are different views. For example, Lewis [52] adopts the modal view but not the subjunctive view. For more on subjunctivism, see Comesaña [15].

Epistemological Points The extent to which subjunctivist-flavored theories of knowledge preserve closure has recently been a topic of active discussion (see, e.g., Alspector-Kelly [3]; Adams et al. [1]). I show (in Section 5) that in contrast to Lewis's (non-subjunctive) theory, the other RA, tracking, and safety theories cited suffer from essentially the same widespread closure failures, far beyond the failure of K, which few if any proponents of these theories would welcome.[4] The theories' structural features responsible for these closure failures also lead (in Section 8) to serious problems of higher-order knowledge, including the possibility of knowing Fitch-paradoxical propositions [23].

Analysis of these results reveals (in Section 9) that two parameters of a modal theory of knowledge affect whether it preserves closure. Each parameter has two values, for four possible parameter settings with respect to which each theory can be classified (Table 2). Of the theories mentioned, only Lewis's, with its unique parameter setting, fully preserves closure (for a fixed context). (In Section 8 I clarify an issue, raised by Williamson [79, 80], about whether Lewis's theory also validates strong principles of higher-order knowledge).

In the terminology of Dretske [19], the knowledge operator for Lewis's theory is *fully penetrating*. For all of the other theories, the knowledge operator lacks the basic closure properties that Dretske wanted from a *semi-penetrating* operator. Contrary to common assumptions in the literature (perhaps due to neglect of the second theory parameter in Section 9), serious closure failures are not avoided by modified subjunctivist theories, such as DeRose's [17] modified tracking theory or the modified safety theory with *bases*, treated formally in Holliday [38, Sections 2.10.1, 2.D]. For those seeking a balance of closure properties between full closure and not enough closure, it appears necessary to abandon essential elements of the standard theories. I show how to do so in part II.

While I take the results of this paper to be negative for subjunctivist-flavored theories qua theories of knowledge, we can also take them to be neutral results about other desirable epistemic properties, viz., the properties of having ruled out the relevant alternatives to a proposition, of having a belief that tracks the truth of a proposition, of having a safe belief in a proposition, etc., even if these are neither necessary nor sufficient for knowledge (see Sections 5 and 7).

Logical Points This paper demonstrates the effectiveness of an alternative approach to proving modal completeness theorems, illustrated by van Benthem [8, Section 4.3] for the normal modal logic **K**, in a case that presents difficulties for a standard canonical model construction. The key element of the alternative approach is a "modal decomposition" result. By proving such results (Theorem 5.2), we will obtain

[4] While closure failures for these subjunctivist-flavored theories go too far in some directions, in other directions they do not go far enough for the purposes of Dretske and Nozick: all of these theories validate closure principles (see Section 5) that appear about as dangerous as K in arguments for radical skepticism about knowledge. This fact undermines the force of responding to skepticism by rejecting K on subjunctivist grounds, as Nozick does.

completeness (Corollary 7.1) of two non-normal modal logics with respect to new semantics mixing elements of ordering semantics [50] and relational semantics [43]. One of these logics, dubbed *the logic of ranked relevant alternatives*, appears not to have been previously identified in the modal logic literature. Further results on decidability (Corollary 5.9), finite models (Corollary 5.24), and complexity (Corollary 5.25) follow from the proof of the modal decomposition results.

In addition to these technical points, the paper aims to show that for modal logicians, epistemology represents an area of sophisticated theorizing in which modal-logical tools can help to clarify and systematize parts of the philosophical landscape. Doing so also benefits modal logic by broadening its scope, bringing interesting new structures and systems under its purview.

In Section 2, I begin with our running example, motivating the issue of epistemic closure. I then introduce the formal framework for the study of closure in RA and subjunctivist theories in Sections 3 and 4. With this setup, I state and prove the main theorems in Sections 5 and 7, with an interlude on relations between RA and subjunctivist models in Section 6. Finally, I investigate higher-order knowledge in Section 8 and discuss the relation between theory parameters and closure failures in Section 9.

Throughout the paper, comments on the faithfulness of the formalization to the philosophical ideas are often in order. To avoid disrupting the flow of presentation, I place some of these important comments in footnotes. Readers who wish to focus on logical ideas should be able to step from definitions to lemmas to theorems, reading the exposition between steps as necessary.

2 The Question of Closure

Example 2.1 (*Medical Diagnosis*) Two medical students, A and B, are subjected to a test. Their professor introduces them to the same patient, who presents various symptoms, and the students are to make a diagnosis of the patient's condition. After some independent investigation, both students conclude that the patient has a common condition c. In fact, they are both correct. Yet only student A passes the test. For the professor wished to see if the students would check for another common condition c' that causes the same visible symptoms as c. While A ran laboratory tests to rule out c' before making the diagnosis of c, B made the diagnosis of c after only a physical exam.

In evaluating the students, the professor concludes that although both gave the correct diagnosis of c, student B did not know that the patient's condition was c, since B did not rule out the alternative of c'. Had the patient's condition been c', student B would (or at least might) still have thought it was c, since the physical exam would not have revealed a difference. Student B was *lucky*. The condition that B associated with the patient's visible symptoms happened to be what the patient had, but if the professor had chosen a patient with c', student B might have made a misdiagnosis. By contrast, student A secured against this possibility of error by running the lab tests. For this reason, the professor judges that student A knew the patient's condition, passing the test.

Of course, $\mathtt{A}$ did not secure against *every* possibility of error. Suppose there is an extremely rare disease[5] x such that people with x appear to have c on lab tests given for c and c', even though people with x are *immune* to c, and only extensive further testing can detect x in its early stages. Should we say that $\mathtt{A}$ did not know that the patient had c after all, since $\mathtt{A}$ did not rule out x? According to a classic *relevant alternatives* style answer (see Goldman [26, 775]; Dretske [21, 365]), the requirement that one rule out *all* possibilities of error would make knowledge impossible, since there are always some possibilities of error—however remote and far-fetched—that are not eliminated by one's evidence and experience. Yet if no one had any special reason to think that the patient may have had x instead of c, then it should not have been necessary to rule out such a remote possibility in order to know that the patient has the common condition (cf. Austin [5, 156ff]; Stroud [71, 51ff]).[6]

If one accepts the foregoing reasoning, then one is close to denying closure under known implication (K). For suppose that student $\mathtt{A}$ knows that if her patient has c, then he does not have x (because x confers immunity to c), (i) $K\,(c \to \neg x)$.[7] Since $\mathtt{A}$ did not run any of the tests that could detect the presence or absence of x, arguably she does not know that the patient does not have x, (ii) $\neg K \neg x$. Given the professor's judgment that $\mathtt{A}$ knows that the patient has condition c, (iii) Kc, together (i)–(iii) violate the following instance of K: (iv) $(Kc \wedge K\,(c \to \neg x)) \to K \neg x$. To retain K, one must say either that $\mathtt{A}$ does not know that the patient has condition c after all (having not excluded x), or else that $\mathtt{A}$ can know that a patient does not have a disease x without running any of the specialized tests for the disease (having learned instead that the patient has c, but from lab results consistent with x).[8] While the second option threatens to commit us to problematic "easy knowledge" [14], the first option threatens to commit us to radical skepticism about knowledge, given the inevitability of uneliminated possibilities of error noted above.

Dretske [19] and Nozick [58] propose to resolve the inconsistency of (i)–(iv), a version of the now standard "skeptical paradox" [13, 17], by denying the validity of K and its instance (iv) in particular. This denial has nothing to do with the "putting two and two together" problem noted in Section 1. The claim is that K would fail even for Dretske's [19] "ideally astute logicians" (1010). I will cash out this phrase as follows: first, such an agent knows all (classically) valid logical principles (*validity omniscience*);[9] second, such an agent believes all the (classical) logical consequences

[5] Perhaps it has never been documented, but it is a possibility of medical theory.

[6] Local skeptics about medical knowledge may substitute one of the standard cases with a similar structure involving, e.g., disguised mules, trick lighting, etc. (see Dretske [19]).

[7] For convenience, I use 'c', 'c'', and 'x' not only as names of medical conditions, but also as symbols for atomic sentences with the obvious intended meanings—that the patient has condition c, c', and x, respectively. Also for convenience, I will not bother to add quotes when mentioning symbolic expressions.

[8] This statement of the dilemma ignores the option of *contextualism*, investigated in Holliday [37, 38]. Stine [70], Lewis [52], and Cohen [13] propose contextualist versions of the RA theory, while DeRose [17] proposes a contextualist version of Nozick's tracking theory. See DeRose [18] for a state of the art treatment of contextualism.

[9] Note the distinction with a stronger property of *consequence omniscience* (standardly "logical omniscience"), that one knows all the logical consequences of what one knows.

of the set of propositions she believes (*full doxastic closure*).[10] Dretske's explanation for why K fails even for such agents is in terms of the RA theory. (We turn to Nozick's view in Section 4). For this theory, to know p is (to truly believe p and) to have *ruled out the relevant alternatives to p*. In coming to know c and $c \to \neg x$, the agent rules out certain relevant alternatives. In order to know $\neg x$, the agent must rule out certain relevant alternatives. But the relevant alternatives in the two cases *are not the same*. According to our earlier reasoning, x is not an alternative that must be ruled out in order for Kc to hold. But x *is* an alternative that must be ruled out in order for $K\neg x$ to hold (cf. Remark 3.9 in Section 3). It is because the relevant alternatives may be different for what is in the antecedent and the consequent of K that instances like (iv) can fail.

In an influential objection to Dretske, Stine [70] claimed that to allow for the relevant alternatives to be different for the premises and conclusion of an argument about knowledge "would be to commit some logical sin akin to equivocation" (256). Yet as Heller [34] points out in Dretske's defence, a similar charge of equivocation could be made (incorrectly) against accepted counterexamples to the principles of transitivity or antecedent strengthening for counterfactuals. If we take a counterfactual $\varphi \mathrel{\Box\!\!\rightarrow} \psi$ to be true iff the "closest" φ-worlds are ψ-worlds, then the inference from $\varphi \mathrel{\Box\!\!\rightarrow} \psi$ to $(\varphi \wedge \chi) \mathrel{\Box\!\!\rightarrow} \psi$ is invalid because the closest $(\varphi \wedge \chi)$-worlds may not be among the closest φ-worlds. Heller argues that there is no equivocation in such counterexamples since we use the same, fixed similarity ordering of worlds to evaluate the different conditionals. Similarly, in the example of closure failure, the most relevant $\neg c$-worlds may differ from the most relevant x-worlds—so one can rule out the former without ruling out the latter—even assuming a fixed relevance ordering of worlds. In this defense of Dretske, Heller brings the RA theory closer to subjunctivist theories that place counterfactual conditions on knowledge.

With this background, let us formulate the question of closure to be studied. We begin with the official definition of our (first) propositional epistemic language. The framework of Sections 3 and 4 could be extended for quantified epistemic logic, but there is already plenty to investigate in the propositional case.[11]

Definition 2.2 (**Epistemic Language**) Let $\mathsf{At} = \{p, q, r, \dots\}$ be a countably infinite set of atomic sentences. The *epistemic language* is defined inductively by

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K\varphi,$$

[10] We may add that such an agent has come to believe these logical consequences by "competent deduction," rather than (only) by some other means, but we will not explicitly represent methods or bases of beliefs here (see Remark 2.3). By "all the logical consequences" I mean all of those *involving concepts that the agent grasps*. Otherwise one might believe p and yet fail to believe $p \vee q$ because one does not grasp q (see Williamson [78, 283]). Assume that the agent grasps all of the atomic $p, q, r, \dots$ of Definition 2.2.

[11] It is not difficult to extend the framework of Sections 3 and 4 to study closure principles of the form shown below where the φ's and ψ's may contain first-order quantifiers, provided that no free variables are allowed within the scope of any K operator. The closure behavior of K with respect to $\forall$ and $\exists$ can be anticipated from the closure behavior of K with respect to $\wedge$ and $\vee$ shown in Theorem 5.2. Of course, interesting complications arise whenever we allow quantification into the scope of a K operator (see Holliday and Perry [41]).

where $p \in \mathsf{At}$. As usual, expressions containing $\vee$, $\rightarrow$, and $\leftrightarrow$ are abbreviations, and by convention $\wedge$ and $\vee$ bind more strongly than $\rightarrow$ or $\leftrightarrow$ in the absence of parentheses; we take $\top$ to be an arbitrary tautology (e.g., $p \vee \neg p$), and $\bot$ to be $\neg\top$. The *modal depth* of a formula φ is defined recursively as follows: $d(p) = 0$, $d(\neg\varphi) = d(\varphi)$, $d(\varphi \wedge \psi) = \max(d(\varphi), d(\psi))$, and $d(K\varphi) = d(\varphi) + 1$. A formula φ is *propositional* iff $d(\varphi) = 0$ and *flat* iff $d(\varphi) \leq 1$.

The flat fragment of the epistemic language has a special place in the study of closure, which need not involve higher-order knowledge. In the most basic case we are interested in whether for a valid propositional formula $\varphi_1 \wedge \cdots \wedge \varphi_n \rightarrow \psi$, the associated "closure principle" $K\varphi_1 \wedge \cdots \wedge K\varphi_n \rightarrow K\psi$ is valid, according to some semantics for the K operator. More generally, we will consider closure principles of the form $K\varphi_1 \wedge \cdots \wedge K\varphi_n \rightarrow K\psi_1 \vee \cdots \vee K\psi_m$, allowing each φ_i and ψ_j to be of arbitrary modal depth. As above, we ask whether such principles hold for ideally astute logicians. The question can be understood in several ways, depending on whether we have in mind what may be called *pure*, *empirical*, or *deductive* closure principles.

Remark 2.3 (*Types of Closure*) For example, if we understand the principle $K(\varphi \wedge \psi) \rightarrow K\psi$ as a *pure* closure principle, then its validity means that an agent cannot know $\varphi \wedge \psi$ without knowing ψ—regardless of whether the agent came to believe ψ by "competent deduction" from $\varphi \wedge \psi$.[12] (Perhaps she came to believe ψ from perception, φ from testimony, and $\varphi \wedge \psi$ by competent deduction from φ and ψ.) More generally, if we understand $K\varphi_1 \wedge \cdots \wedge K\varphi_n \rightarrow K\psi$ as a pure closure principle, its validity means that an agent cannot know $\varphi_1, \ldots, \varphi_n$ without knowing ψ. Understood as an *empirical* closure principle, its validity means that an agent who has done enough empirical investigation to know $\varphi_1, \ldots, \varphi_n$ has done enough to know ψ. Finally, understood as a *deductive* closure principle, its validity means that *if* the agent came to believe ψ from $\varphi_1, \ldots, \varphi_n$ by competent deduction, all the while knowing $\varphi_1, \ldots, \varphi_n$, then she knows ψ. As suggested by Williamson [78, 282f], it is highly plausible that $K(\varphi \wedge \psi) \rightarrow K\psi$ is a pure (and hence empirical and deductive) closure principle. By contrast, closure under known implication is typically understood as only an empirical or deductive closure principle.[13] Here we will not explicitly represent in our language or models the idea of deductive closure. I do so elsewhere [38, Section 2.D] in formalizing versions of the tracking and safety theories that take into account *methods* or *bases* of beliefs. It is first necessary to

[12] Harman and Sherman [31] criticize Williamson's [78] talk of "deduction" as extending knowledge for its "presupposition that deduction is a kind of inference, something one does" (495). Our talk of an agent coming to believe ψ by "competent deduction" from $\varphi_1, \ldots, \varphi_n$ can be taken as elliptical for the following (cf. Harman [31, 496]): the agent constructs a valid deduction from believed premises $\varphi_1, \ldots, \varphi_n$ to conclusion ψ, recognizes that the construction is a valid deduction, and comes to believe ψ on that basis.

[13] Deductive closure principles belong to a more general category of "active" closure principles, which are conditional on the agent performing some action, of which deduction is one example. As Johan van Benthem (personal communication) suggests, the active analogue of K has the form $K\varphi \wedge K(\varphi \rightarrow \psi) \rightarrow [a]K\psi$, where $[a]$ stands for *after action a*.

understand the structural reasons for why the basic RA, tracking, and safety conditions are not purely or empirically closed, in order to understand whether the refined theories solve all the problems of epistemic closure.[14]

3 Relevant Alternatives

In this section, I introduce formalizations of two RA theories of knowledge. Before giving RA semantics for the epistemic language of Definition 2.2, let us observe several distinctions between different versions of the RA theory.

The first concerns the nature of the "alternatives" that one must rule out to know p. Are they *possibilities* (or ways the world could/might be) in which p is false?[15] Or are they *propositions* incompatible with p? Both views are common in the literature, sometimes within a single author. Although earlier I wrote in a way suggestive of the second view, in what follows I adopt the first view, familiar in the epistemic logic tradition since Hintikka, since it fits the theories I will formalize. For a comparison of the views, see Holliday [38, Section 4.A].

The second distinction concerns the structure of relevant alternatives. On one hand, Dretske [21] states the following definition in developing his RA theory: "call the set of possible alternatives that a person must be in an evidential position to exclude (when he knows *P*) the *Relevancy Set* (RS)" (371). On the other hand, Heller [34] considers (and rejects) an interpretation of the RA theory in which "there is a certain set of worlds selected as relevant, and S must be able to rule out the not-p worlds within that set" (197).

According to Dretske, for every proposition P, there is a relevancy set for that P. Let us translate this into Heller's talk of worlds. Where $\overline{P}$ is the set of all worlds in which P is false, let $\mathsf{r}(P)$ be the relevancy set for P, so $\mathsf{r}(P) \subseteq \overline{P}$. To be more precise, since objective features of an agent's situation in world w may affect what alternatives are relevant and therefore what it takes to know P in w (see Dretske [21, 377] and DeRose [18, 30f] on "subject factors"), let us write '$\mathsf{r}(P, w)$' for the relevancy set for P in world w, so $\mathsf{r}(P, w)$ may differ from $\mathsf{r}(P, v)$ for a distinct world v in which the agent's situation is different. Finally, if we allow (unlike Dretske) that the conversational context $\mathcal{C}$ of those attributing knowledge to the agent can also affect what alternatives are relevant in a given situation w and therefore what it takes to count as knowing P in w relative to $\mathcal{C}$ (see [18, 30f] on "attributor factors"), then we should write '$\mathsf{r}_{\mathcal{C}}(P, w)$' to make the relativization to context explicit.

[14] There are problematic failures of pure and deductive closure for the tracking theory *with methods*, for the structural reasons identified here. The safety theory *with bases* may support deductive closure (although see Alspector-Kelly [3]), but it also has problems with pure closure for the structural reasons identified here. See Holliday [38, Section 2.D].

[15] In order to deal with self-locating knowledge, one may take the alternatives to be "centered" worlds or possible individuals (see Lewis [51, Section 1.4] and references therein). Another question is whether we should think of what is ruled out by knowledge as including *ways the world could not be* (metaphysically "impossible worlds" or even logically impossible worlds), in addition to *ways the world could be*. See King [42] on this question and Chalmers [11] on *ways the world might be* vs. *ways the world might have been*.

The quote from Dretske suggests the following definition:

> According to a $\mathsf{RS}_{\forall\exists}$ theory, for every context $\mathcal{C}$, for every world w, and for every ($\forall$) proposition P, there is ($\exists$) a set of *relevant* (*in* w) *not-P worlds*, $\mathsf{r}_{\mathcal{C}}(P, w) \subseteq \overline{P}$, such that in order to know P in w (relative to $\mathcal{C}$) one must rule out the worlds in $\mathsf{r}_{\mathcal{C}}(P, w)$.

By contrast, the quote from Heller suggests the following definition:

> According to a $\mathsf{RS}_{\exists\forall}$ theory, for every context $\mathcal{C}$ and for every world w, there is ($\exists$) a set of *relevant* (*in* w) *worlds*, $\mathsf{R}_{\mathcal{C}}(w)$, such that for every ($\forall$) proposition P, in order to know P in w (relative to $\mathcal{C}$) one must rule out the not-P worlds in that set, i.e., the worlds in $\mathsf{R}_{\mathcal{C}}(w) \cap \overline{P}$.

As a simple logical observation, every $\mathsf{RS}_{\exists\forall}$ theory is a $\mathsf{RS}_{\forall\exists}$ theory (take $\mathsf{r}_{\mathcal{C}}(P, w) = \mathsf{R}_{\mathcal{C}}(w) \cap \overline{P}$), but not necessarily vice versa. From now on, when I refer to $\mathsf{RS}_{\forall\exists}$ theories, I have in mind theories that are not also $\mathsf{RS}_{\exists\forall}$ theories. This distinction is at the heart of the disagreement about epistemic closure between Dretske and Lewis [52], as Lewis clearly adopts an $\mathsf{RS}_{\exists\forall}$ theory.

In a *contextualist* $\mathsf{RS}_{\exists\forall}$ theory, such as Lewis's, the set of relevant worlds may change as context changes. Still, for any given context $\mathcal{C}$, there is a set $\mathsf{R}_{\mathcal{C}}(w)$ of relevant (at w) worlds, which does not depend on the particular proposition in question. The $\mathsf{RS}_{\forall\exists}$ vs. $\mathsf{RS}_{\exists\forall}$ distinction is about how theories view the relevant alternatives *with respect to a fixed context*. Here we study which closure principles hold for different theories with respect to a fixed context. Elsewhere I extend the framework to context change [37, 38].

A third distinction between versions of the RA theory concerns different notions of ruling out or eliminating alternatives (possibilities or propositions). On one hand, Lewis [52] proposes that "a possibility … [v] … is *uneliminated* iff the subject's perceptual experience and memory in … [v] … exactly match his perceptual experience and memory in actuality" (553). On the other hand, Heller [34] proposes that "S's ability to rule out not-p be understood thus: S does not believe p in any of the relevant not-p worlds" (98). First, we model the RA theory with a Lewis-style notion of elimination. By 'Lewis-style', I do not mean a notion that involves experience or memory; I mean any notion of elimination that allows us to decide whether a possibility v is eliminated by an agent in w *independently* of any proposition P under consideration, as Lewis's notion does. In Section 4, we turn to Heller's notion, which is closely related to Nozick's [58] tracking theory. We compare the two notions in Section 9.

Below we define our first class of models, following Heller's RA picture of "worlds surrounding the actual world ordered according to how realistic they are, so that those worlds that are more realistic are closer to the actual world than the less realistic ones" [33, 25] with "those that are too far away from the actual world being irrelevant" [34, 199]. These models represent the epistemic state of an agent from a third-person perspective. We should not assume that anything in the model is something that the agent has in mind. Contextualists should think of the model $\mathcal{M}$ as associated with a fixed context of knowledge attribution, so a change in context corresponds to a change in models from $\mathcal{M}$ to $\mathcal{M}'$ (an idea formalized in Holliday [37, 38]). Just as the model is not something that the agent has in mind, it is not something that particular speakers attributing knowledge to the agent have in mind either. For possibilities may be relevant and hence should be included in our model, even if the attributors are not considering them (see DeRose [18, 33]).

Finally, for simplicity (and in line with Lewis [52]) we will not represent in our RA models an agent's beliefs separately from her knowledge. Adding the doxastic machinery of Section 4 (which guarantees doxastic closure) would be easy, but if the only point were to add *believing* φ as a necessary condition for knowing φ, this would not change any of our results about RA knowledge.[16]

Definition 3.1 (**RA Model**) A *relevant alternatives model* is a tuple $\mathcal{M}$ of the form $\langle W, \rightarrowtriangle, \preceq, V\rangle$ where:

1. W is a nonempty set;
2. $\rightarrowtriangle$ is a reflexive binary relation on W;
3. $\preceq$ assigns to each $w \in W$ a binary relation $\preceq_w$ on some $W_w \subseteq W$;
   (a) $\preceq_w$ is reflexive and transitive;
   (b) $w \in W_w$, and for all $v \in W_w$, $w \preceq_w v$;
4. V assigns to each $p \in \mathsf{At}$ a set $V(p) \subseteq W$.

For $w \in W$, the pair $\mathcal{M}, w$ is a *pointed* model.

I refer to elements of W as "worlds" or "possibilities" interchangeably.[17] As usual, think of $V(p)$ as the set of worlds where the atomic sentence p holds.

Take $w \rightarrowtriangle v$ to mean that v is an *uneliminated* possibility for the agent in w.[18] For generality, I assume only that $\rightarrowtriangle$ is reflexive, reflecting the fact that an agent can never eliminate her actual world as a possibility. According to Lewis's [52] notion of elimination, $\rightarrowtriangle$ is an equivalence relation. However, whether we assume transitivity and symmetry in addition to reflexivity does not affect our main results (see Remark 5.20). This choice only matters if we make further assumptions about the $\preceq_w$ relations, discussed in Section 8.

[16] If one were to also adopt a variant of Lewis's [52] *Rule of Belief* according to which any world v doxastically accessible for the agent in w must be relevant and uneliminated for the agent in w (i.e., using notation introduced below, wDv implies $v \in \mathrm{Min}_{\preceq_w}(W)$ and $w \rightarrowtriangle v$), then belief would already follow from the knowledge condition of Definition 3.4.

[17] Lewis [52] is neutral on whether the *possibilities* referred to in his definition of knowledge must be "maximally specific" (552), as *worlds* are often thought to be. It should be clear that our examples do not depend on taking possibilities to be maximally specific either.

[18] Those who have used standard Kripke models for epistemic modeling should note an important difference in how we use W and $\rightarrowtriangle$. We include in W not only possibilities that the agent has not eliminated, but also possibilities that the agent *has* eliminated, including possibilities v such that $w \not\rightarrowtriangle v$ for all w distinct from v. While in standard Kripke semantics for the (single-agent) epistemic language, such a possibility v can always be deleted from W without changing the truth value of any formula at w (given the invariance of truth under $\rightarrowtriangle$-generated submodels), this will *not* be the case for one of our semantics below (D-semantics). So if we want to indicate that an agent in w has eliminated a possibility v, we do not leave it out of W; instead, we include it in W and set $w \not\rightarrowtriangle v$.

Take $u \preceq_w v$ to mean that u is *at least as relevant* (at w) as v is.[19] A relation satisfying Definition 3.1.3a is a *preorder*. The family of preorders in an RA model is like one of Lewis's (weakly centered) comparative similarity systems [49, Section 2.3] or standard γ-models [48], but without his assumption that each $\preceq_w$ is *total* on its field W_w (see Def. 3.3.3). Condition 3b, that w is at least as relevant at w as any other world is, corresponds to Lewis's [52] *Rule of Actuality*, that "actuality is always a relevant alternative" (554).

By allowing $\preceq_w$ and $\preceq_v$ to be different for distinct worlds w and v, we allow the world-relativity of comparative relevance (based on differences in "subject factors") discussed above. A fixed context may help to determine not only which possibilities are relevant, given the way things actually are, but also which possibilities would be relevant were things different. Importantly, we also allow $\preceq_w$ and $\preceq_v$ to be different when v is an uneliminated possibility for the agent in w, so $w \rightarrowtriangle v$. In other words, we do not assume that in w the agent can eliminate any v for which $\preceq_v \neq \preceq_w$. As Lewis [52] put it, "the subject himself may not be able to tell what is properly ignored" (554). We will return to these points in Section 8 in our discussion of higher-order knowledge.

Notation 3.2 (**Derived Relations, Min**) Where $w, v, u \in W$ and $S \subseteq W$,

- $u \prec_w v$ iff $u \preceq_w v$ and not $v \preceq_w u$; and $u \simeq_w v$ iff $u \preceq_w v$ and $v \preceq_w u$;
- $\text{Min}_{\preceq_w}(S) = \{v \in S \cap W_w \mid \text{there is no } u \in S \text{ such that } u \prec_w v\}$.

Hence $u \prec_w v$ means that possibility u is *more relevant* (at w) than possibility v is, while $u \simeq_w v$ means that they are equally relevant. $\text{Min}_{\preceq_w}(S)$ is the set of *most relevant* (at w) possibilities out of those in S that are ordered by $\preceq_w$, in the sense that there are no other possibilities that are more relevant (at w).

Definition 3.3 (**Types of Orderings**) Consider an RA model $\mathcal{M} = \langle W, \rightarrowtriangle, \preceq, V\rangle$ with $w \in W$.

1. $\preceq_w$ is *well-founded* iff for every nonempty $S \subseteq W_w$, $\text{Min}_{\preceq_w}(S) \neq \emptyset$;
2. $\preceq_w$ is *linear* iff for all $u, v \in W_w$, either $u \prec_w v$, $v \prec_w u$, or $u = v$;
3. $\preceq_w$ is *total* iff for all $u, v \in W_w$, $u \preceq_w v$ or $v \preceq_w u$;
4. $\preceq_w$ has a *universal field* iff $W_w = W$;
5. $\preceq_w$ is *centered* (*weakly centered*) iff $\text{Min}_{\preceq_w}(W) = \{w\}$ ($w \in \text{Min}_{\preceq_w}(W)$).

If a property holds of $\preceq_v$ for all $v \in W$, then we say that $\mathcal{M}$ has the property.

Well-foundedness is a (language-independent) version of the "Limit Assumption" discussed by Lewis [49, Section 1.4]. Together well-foundedness and linearity amount to "Stalnaker's Assumption" (ibid., Section 3.4). Totality says that any worlds in the field of $\preceq_w$ are comparable in relevance. So a total preorder $\preceq_w$ is a relevance

[19] One might expect $u \preceq_w v$ to mean that v is at least as relevant (at w) as u is, by analogy with $x \leq y$ in arithmetic, but Lewis's [49, Section 2.3] convention is now standard.

ranking of worlds in W_w. Universality (ibid., Section 5.1) says that all worlds are assessed for relevance at w. Finally, (with Def. 3.1.3b) centering (ibid., Section 1.3) says that w is *the* most relevant world at w, while weak centering (ibid., Section 1.7) (implied by Def. 3.1.3b) says that w is *among* the most relevant.

I assume well-foundedness (always satisfied in finite models) in what follows, since it allows us to state more perspicuous truth definitions. However, this assumption does not affect our results (see Remark 5.13). By contrast, totality does make a difference in valid closure principles for one of our theories (see Fact 5.7), while the addition of universality does not (see Prop. 5.23). I comment on linearity and centering vs. weak centering after Definition 3.6.

We now interpret the epistemic language of Definition 2.2 in RA models, considering three semantics for the K operator. I call these C-semantics, for **C**artesian, D-semantics, for **D**retske, and L-semantics, for **L**ewis. C-semantics is not intended to capture Descartes' view of knowledge. Rather, it is supposed to reflect a high standard for the truth of knowledge claims—knowledge requires ruling out all possibilities of error, however remote—in the spirit of Descartes' worries about error in the First Meditation; formally, C-semantics is just the standard semantics for epistemic logic in the tradition of Hintikka [36], but I reserve 'H-semantics' for later. D-semantics is one way (but not the only way) of understanding Dretske's [21] $\mathsf{RS}_{\forall\exists}$ theory, using Heller's [33, 34] picture of relevance orderings over possibilities.[20] Finally, L-semantics follows Lewis's [52] $\mathsf{RS}_{\exists\forall}$ theory (for a fixed context).

Definition 3.4 (**Truth in an RA Model**) Given a well-founded RA model $\mathcal{M} = \langle W, \rightarrowtriangle, \preceq, V\rangle$ with $w \in W$ and a formula φ in the epistemic language, define $\mathcal{M}, w \vDash_x \varphi$ (φ is true at w in $\mathcal{M}$ according to X-semantics) as follows:

$$\begin{array}{ll}
\mathcal{M}, w \vDash_x p & \text{iff } w \in V(p);\\
\mathcal{M}, w \vDash_x \neg\varphi & \text{iff } \mathcal{M}, w \nvDash_x \varphi;\\
\mathcal{M}, w \vDash_x (\varphi \wedge \psi) & \text{iff } \mathcal{M}, w \vDash_x \varphi \text{ and } \mathcal{M}, w \vDash_x \psi.
\end{array}$$

For the K operator, the C-semantics clause is that of standard modal logic:

$$\mathcal{M}, w \vDash_c K\varphi \text{ iff } \forall v \in W\colon \text{if } w \rightarrowtriangle v \text{ then } \mathcal{M}, v \vDash_c \varphi,$$

which states that φ is known at w iff φ is true in all possibilities uneliminated at w. I will write this clause in another, equivalent way below, for comparison with the D- and L-semantics clauses. First, we need two pieces of notation.

Notation 3.5 (**Extension and Complement**) Where $\mathcal{M} = \langle W, \rightarrowtriangle, \preceq, V\rangle$,

- $[\![\varphi]\!]^{\mathcal{M}}_x = \{v \in W \mid \mathcal{M}, v \vDash_x \varphi\}$ is the set of worlds where φ is true in $\mathcal{M}$ according to X-semantics; if $\mathcal{M}$ and x are clear from context, I write '$[\![\varphi]\!]$'.

[20] In part II, I argue that there is a better way of understanding Dretske's [21] $\mathsf{RS}_{\forall\exists}$ theory, without the familiar world-ordering picture. Hence I take the 'D' for D-semantics as loosely as the 'C' for C-semantics. Nonetheless, it is a helpful mnemonic for remembering that D-semantics formalizes an RA theory that allows closure failure, as Dretske's does, while L-semantics formalizes an RA theory that does not, like Lewis's.

- For $S \subseteq W$, $\overline{S} = \{v \in W \mid v \notin S\}$ is the complement of S in W. When W may not be clear from context, I write '$W \setminus S$' instead of '$\overline{S}$'.

Definition 3.6 (Truth in an RA Model cont.) For C-, D-, and L-semantics, the clauses for the K operator are:[21]

$$\begin{aligned}
&\mathcal{M}, w \vDash_c K\varphi \ \text{ iff } \forall v \in \overline{[\![\varphi]\!]_c}\colon w \not\rightarrowtriangle v;\\
&\mathcal{M}, w \vDash_d K\varphi \ \text{ iff } \forall v \in \mathrm{Min}_{\preceq_w}\left(\overline{[\![\varphi]\!]_d}\right) : w \not\rightarrowtriangle v;\\
&\mathcal{M}, w \vDash_l K\varphi \ \text{ iff } \forall v \in \ \mathrm{Min}_{\preceq_w}(W) \cap \overline{[\![\varphi]\!]_l}\colon w \not\rightarrowtriangle v.
\end{aligned}$$

According to C-semantics, in order for an agent to know φ in world w, *all* of the $\neg\varphi$-possibilities must be eliminated by the agent in w. According to D-semantics, for any φ there is a set $\mathrm{Min}_{\preceq_w}\left(\overline{[\![\varphi]\!]_d}\right)$ of *most relevant* (*at* w) $\neg\varphi$-possibilities that the agent must eliminate in order to know φ. Finally, according to L-semantics, there is a set of relevant possibilities, $\mathrm{Min}_{\preceq_w}(W)$, such that for any φ, in order to know φ the agent must eliminate the $\neg\varphi$-possibilities *within that set.* Recall the $\mathsf{RS}_{\forall\exists}$ vs. $\mathsf{RS}_{\exists\forall}$ distinction above.

If φ is true at all pointed models according to X-semantics, then φ is *X-valid*, written '$\vDash_x \varphi$'. Since the semantics do not differ with respect to propositional formulas φ, I sometimes omit the subscript in '$\vDash_x$' and simply write '$\mathcal{M}, w \vDash \varphi$'. These conventions also apply to the semantics in Definition 4.3.

Since for L-semantics we think of $\mathrm{Min}_{\preceq_w}(W)$ as the set of simply *relevant* worlds, ignoring the rest of $\preceq_w$, we allow $\mathrm{Min}_{\preceq_w}(W)$ to contain multiple worlds. Hence with L-semantics we assume neither centering nor linearity, which implies centering by Definition 3.1.3b. For D-semantics, whether we assume centering/linearity does not affect our results (as shown in Section 5.2.2).

It is easy to check that according to C/D/L-semantics, whatever is known is true. For D- and L-semantics, Fact 3.7 reflects Lewis's [52, 554] observation that the veridicality of knowledge follows from his Rule of Actuality, given that an agent can never eliminate her actual world as a possibility. Formally, veridicality follows from the fact that w is minimal in $\preceq_w$ and $w \rightarrowtriangle w$.

Fact 3.7 (Veridicality) *$K\varphi \rightarrow \varphi$ is C/D/L-valid.*

Consider the model in Fig. 1, drawn for student A in Example 2.1. An arrow from w to v indicates that $w \rightarrowtriangle v$, i.e., v is uneliminated by the agent in w. (For all $v \in W$, $v \rightarrowtriangle v$, but we omit all reflexive loops). The ordering of the worlds by their relevance

[21] Instead of thinking in terms of three different satisfaction relations, $\vDash_c$, $\vDash_d$, and $\vDash_l$, some readers may prefer to think in terms of one satisfaction relation, $\vDash$, and three different operators, K_c, K_d, and K_l. I choose to subscript the turnstile instead of the operator in order to avoid proliferating subscripts in formulas. One should not read anything more into this practical choice of notation. (However, note that epistemologists typically take themselves to be proposing different accounts of the conditions under which an agent has knowledge, rather than proposing different epistemic notions of knowledge$_1$, knowledge$_2$, etc.)

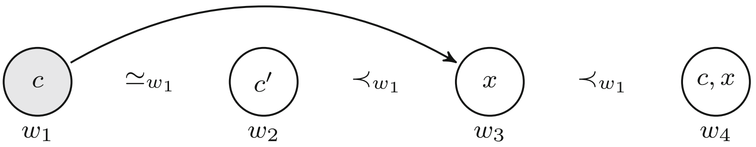

Fig. 1 An RA model for Example 2.1 (partially drawn, reflexive loops omitted)

at w_1, which we take to be the actual world, is indicated between worlds.[22] In w_1, the patient has the common condition *c*, represented by the atomic sentence *c* true at w_1 (see footnote 7). Possibility w_2, in which the patient has the other common condition c' instead of *c*, is just as relevant as w_1. Since the model is for student A, who ran the lab tests to rule out c', A has eliminated w_2 in w_1. A more remote possibility than w_2 is w_3, in which the patient has the rare disease *x*. Since A has not run any tests to rule out *x*, A has not eliminated w_3 in w_1. Finally, the most remote possibility of all is w_4, in which the patient has both *c* and *x*. We assume that A has learned from textbooks that *x* confers immunity to *c*, so A has eliminated w_4 in w_1.

Now consider C-semantics. In discussing Example 2.1, we held that student A knows that the patient's condition is *c*, despite the fact that A did not rule out the remote possibility of the patient's having *x*. C-semantics issues the opposite verdict. According to C-semantics, Kc is true at w_1 iff *all* $\neg c$-worlds, regardless of their relevance, are ruled out by the agent in w_1. However, w_3 is not ruled out by A in w_1, so Kc is false at w_1. Nonetheless, A has some knowledge in w_1. For example, one can check that $K(\neg x \to c)$ is true at w_1.

Remark 3.8 (*Skepticism*) A skeptic might argue, however, that we have failed to include in our model a particular possibility, far-fetched but uneliminated, in which the patient has neither *x* nor *c*, the inclusion of which would make even $K(\neg x \to c)$ false at w_1 according to C-semantics. In this way, C-semantics plays into the hands of skeptics. By contrast, L- and D-semantics help to avoid skepticism by not requiring the elimination of every far-fetched possibility.

Consider the model in Fig. 1 from the perspective of L-semantics. According to L-semantics, student A *does* know that the patient has condition *c*. Kc is true at w_1, because *c* is true in all of the most relevant and uneliminated (at w_1) worlds, namely w_1 itself. Moreover, although A has not ruled out the possibility w_3 in which the patient has disease *x*, according to L-semantics she nonetheless *knows* that the patient does not have *x*. $K\neg x$ is true at w_1, because $\neg x$ is true in all of the most relevant (at w_1) worlds: w_1 and w_2. Indeed, note that $K\neg x$ would be true at w_1 no matter how we defined the $\rightarrowtriangle$ relation.

[22] We ignore the relevance orderings for the other worlds. We also ignore which possibilities are ruled out at worlds other than w_1, since we are not concerned here with student A's higher-order knowledge at w_1. If we were, then we might include other worlds in the model.

Remark 3.9 (*Vacuous Knowledge*) What this example shows is that according to L-semantics, in some cases an agent can know some φ *with no requirement of ruling out possibilities*, i.e., with no requirement on $\rightarrowtriangle$, simply because none of the accessible $\neg\varphi$-possibilities are relevant at w, i.e., because they are not in $\text{Min}_{\preceq_w}(W)$. This is the position of Stine [70, 257] and Rysiew [64, 265], who hold that one can know that skeptical hypotheses do not obtain, without any evidence, simply because the skeptical possibilities are not relevant in the context (also see Lewis [52, 561f]). In general, on the kind of $\mathsf{RS}_{\exists\forall}$ view represented by L-semantics, an agent can know a *contingent empirical truth* φ with no requirement of empirically eliminating any possibilities. Heller [34, 207] rejects such "vacuous knowledge," and elsewhere I discuss this *problem of vacuous knowledge* at length ([40]; also see Cohen [13, 99]; Vogel [74, 158f]; and Remark 4.6 below). By contrast, on the kind of $\mathsf{RS}_{\forall\exists}$ view represented by D-semantics, as long as there is an accessible $\neg\varphi$-possibility, there will be some most relevant (at w) $\neg\varphi$-possibility that the agent must rule out in order to know φ in w. Hence D-semantics avoids vacuous knowledge.

D-semantics avoids the skepticism of C-semantics and the vacuous knowledge of L-semantics, but at a cost for closure. Consider the model in Fig. 1 from the perspective of D-semantics. First observe that D-semantics issues our original verdict that student $\mathtt{A}$ knows that the patient's condition is c. Kc is true at w_1 since the most relevant (at w_1) $\neg c$-world, w_2, is ruled out by $\mathtt{A}$ in w_1. $K(c \to \neg x)$ is also true at w_1, since the most relevant (at w_1) $\neg(c \to \neg x)$-world, w_4, is ruled out by $\mathtt{A}$ in w_1. Not only that, but $K(c \leftrightarrow \neg x)$ is true at w_1, since the most relevant (at w_1) $\neg(c \leftrightarrow \neg x)$-world, w_2, is ruled out by $\mathtt{A}$ in w_1. However, the most relevant (at w_1) x-world, w_3, is *not* ruled out by $\mathtt{A}$ in w_1, so $K\neg x$ is false at w_1. Hence $\mathtt{A}$ does not know that the patient does not have disease x. We have just established the second part of the following fact, which matches Dretske's [19] view. The first part, which follows directly from the truth definition, matches Lewis's [52, 563n21] view.

Fact 3.10 (**Known Implication**) *The principles*

$$K\varphi \wedge K(\varphi \to \psi) \to K\psi \text{ and } K\varphi \wedge K(\varphi \leftrightarrow \psi) \to K\psi$$

are C/L-valid, but not D-valid.[23]

In Dretske's [19, 1007] terminology, Fact 3.10 shows that the knowledge operator in D-semantics is not *fully penetrating*, since it does not penetrate to all of the logical consequence of what is known. Yet Dretske claims that the knowledge operator is *semi-penetrating*, since it does penetrate to some logical consequences: "it seems to me fairly obvious that if someone knows that P and Q, he thereby knows that Q" and "If he knows that P is the case, he knows that P or Q is the case" (1009). This is supposed to be the "trivial side" of Dretske's thesis (ibid.). However, if we understand the RA theory according to D-semantics, then even these monotonicity principles

[23] It is easy to see that for D-semantics (and H/N/S-semantics in Section 4), knowledge fails to be closed not only under known material implication, but even under known *strict* implication: $K\varphi \wedge K\Box(\varphi \to \psi) \to K\psi$, with the $\Box$ in Definition 8.5 (or even the universal modality).

fail (as they famously do for Nozick's theory, discussed in Section 4, for the same structural reasons).

Fact 3.11 (**Distribution and Addition**) *The principles*

$$K\,(\varphi \wedge \psi) \to K\varphi \wedge K\psi \text{ and } K\varphi \to K\,(\varphi \vee \psi)$$

are C/L-valid, but not D-valid.

Proof The proof of C/L-validity is routine. For D-semantics, the pointed model $\mathcal{M}, w_1$ in Fig. 1 falsifies $K(c \wedge \neg x) \to K\neg x$ and $Kc \to K(c \vee \neg x)$. These principle are of the form $K\alpha \to K\beta$. In both cases, the most relevant (at w_1) $\neg\alpha$-world in $\mathcal{M}$ is w_2, which is eliminated by the agent in w_1, so $K\alpha$ is true at w_1. However, in both cases, the most relevant (at w_1) $\neg\beta$-world in $\mathcal{M}$ is w_3, which is uneliminated by the agent in w_1, so $K\beta$ is false at w_1. □

Fact 3.11 is only the tip of the iceberg, the full extent of which is revealed in Section 5. But it already points to a dilemma. On the one hand, if we understand the RA theory according to D-semantics, then the knowledge operator lacks even the basic closure properties that Dretske wanted from a semi-penetrating operator, contrary to the "trivial side" of his thesis; here we have an example of what I called the *problem of containment* in Section 1. On the other hand, if we understand the RA theory according to L-semantics, then the knowledge operator is a fully-penetrating operator, contrary to the non-trivial side of Dretske's thesis; and we have the problem of vacuous knowledge. It is difficult to escape this dilemma while retaining something like Heller's [33, 34] world-ordering picture with which we started before Definition 3.1. However, Dretske's [21] discussion of relevancy sets leaves open whether the RA theory should be developed along the lines of this world-ordering picture. In part II, I will propose a different way of developing the theory so that the knowledge operator is semi-penetrating in Dretske's sense, avoiding the dilemma above.

4 Counterfactuals and Beliefs

In this section, I introduce the formalizations of Heller's [33, 34] RA theory, Nozick's [58] tracking theory, and Sosa's [67] safety theory. Let us begin by defining another class of models, closely related to RA models.

Definition 4.1 (**CB Model**) A *counterfactual belief model* is a tuple $\mathcal{M}$ of the form $\langle W, D, \leqslant, V\rangle$ where W, $\leqslant$, and V are defined in the same way as W, $\preceq$, and V in Definition 3.1, and D is a serial binary relation on W.

Notation 3.2 and Definition 3.3 apply to CB models as for RA models, but with $\leqslant_w$ in place of $\preceq_w$, $<_w$ in place of $\prec_w$, and $\equiv_w$ in place of $\simeq_w$.

Think of D as a *doxastic accessibility* relation, so that wDv indicates that everything the agent believes in w is true in v [51, Section 1.4]. For convenience, we extend the epistemic language of Definition 2.2 to an *epistemic-doxastic* language with a belief operator B for the D relation. We do so in order to state perspicuous truth definitions for the K operator, which could be equivalently stated in a more direct (though cumbersome) way in terms of the D relation. Our main result will be given for the pure epistemic language.

Think of $\leqslant_w$ either as a relevance relation as before (for Heller) or as a relation of *comparative similarity* with respect to world w, used for assessing counterfactuals as in Lewis [49].[24] With the latter interpretation, we can capture the following well-known counterfactual conditions on an agent's belief that φ: if φ were false, the agent would not believe φ (*sensitivity*); if φ were true, the agent would believe φ (*adherence*); the agent would believe φ only if φ were true (*safety*). Nozick [58] argued that sensitivity and adherence—the conjunction of which is *tracking*—are necessary and sufficient for one's belief to constitute knowledge,[25] while Sosa [67] argued that safety is necessary. (In Holliday [38, Section 2.D], I consider the revised tracking and safety theories that take into account methods and bases of belief). Following Nozick and Sosa, we can interpret sensitivity as the counterfactual $\neg\varphi \mathrel{\Box\!\!\rightarrow} \neg B\varphi$, adherence as $\varphi \mathrel{\Box\!\!\rightarrow} B\varphi$, and safety as $B\varphi \mathrel{\Box\!\!\rightarrow} \varphi$, with the caveat in Observation 4.5 below. I will understand the truth of counterfactuals following Lewis [49, 20], such that $\varphi \mathrel{\Box\!\!\rightarrow} \psi$ is true at a world w iff the closest φ-worlds to w according to $\leqslant_w$ are ψ-worlds, subject to the same caveat.[26] The formalization is also compatible with the view that the conditions above should be understood in terms of "close enough" rather than closest worlds. [27]

[24] Heller [33] argues that the orderings for relevance and similarity are the same, only the boundary of the relevant worlds that one must rule out in order to know may extend beyond that of the most similar worlds. See the remarks in note 26 below.

[25] Nozick used the term 'variation' for what I call 'sensitivity' and used 'sensitivity' to cover both variation and adherence; but the narrower use of 'sensitivity' is now standard.

[26] Nozick [58, 680n8] tentatively proposes alternative truth conditions for counterfactuals. However, he also indicates that his theory may be understood in terms of Lewis's semantics for counterfactuals (but see Observation 4.5). This has become the standard practice in the literature. For example, see Vogel [73], Comesaña [15], and Alspector-Kelly [3].

[27] In Definition 4.3, I state the sensitivity, adherence, and safety conditions using the $\mathrm{Min}_{\leqslant_w}$ operator, which when applied to a set S of worlds gives the set of "closest" worlds to w out of those in S. This appears to conflict with the views of Heller [33, 34], who argues for a "close *enough* worlds" analysis rather than a "clos*est* worlds" analysis for sensitivity, and of Pritchard [60, 72], who argues for considering *nearby* rather than only *nearest* worlds for safety and sensitivity. However, the conflict is merely apparent. For if one judges that the clos*est* worlds in a set S, according to $\leqslant_w$, do not include all of the worlds in S that are close *enough*, then we can relax $\leqslant_w$ to a coarser preorder $\leqslant'_w$, so that the closest worlds in S according to $\leqslant'_w$ are exactly those worlds in S previously judged to be closest or close enough. To be precise, given a set $CloseEnough(w) \subseteq W_w$ such that $\mathrm{Min}_{\leqslant_w}(W) \subseteq CloseEnough(w)$ and whenever $y \in CloseEnough(w)$ and $x \leqslant_w y$, then $x \in CloseEnough(w)$, define $\leqslant'_w$ as follows: $v \leqslant'_w u$ iff either $v \leqslant_w u$ or [$u \leqslant_w v$ and $v \in CloseEnough(w)$]. Then $\mathrm{Min}_{\leqslant'_w}(S) = \mathrm{Min}_{\leqslant_w}(S) \cup (CloseEnough(w) \cap S)$, so the close enough S-worlds are included, as desired. For the coarser preorder $\leqslant'_w$, $\mathrm{Min}_{\leqslant'_w}(W) = CloseEnough(w)$ would be the set of worlds close enough/nearby to w. Here we assume, following Heller [34, 201f], that whether a world counts as *close enough/nearby* may be context dependent, but for a fixed context, whether a world is close enough/nearby is not relative to the φ for which we are assessing $K\varphi$ (cf. Cross [16] on counterfactual conditionals and antecedent-relative comparative world similarity); as discussed in Section 2, the fact that (for a given world) there is a single, fixed ordering on the set of worlds is what Heller [34] uses to reply to Stine's [70] equivocation charge against Dretske. Finally, note that while the coarser preorder $\leqslant'_w$ may not be the appropriate relation for assessing counterfactuals, according to the Heller/Pritchard view, it would be appropriate for assessing knowledge.

We are now prepared to define three more semantics for the K operator: H-semantics for **H**eller, N-semantics for **N**ozick, and S-semantics for **S**osa.

Remark 4.2 (*Necessary Conditions*) In defining these semantics, I assume that each theory proposes necessary and sufficient conditions for knowledge. This is true of Nozick's [58] theory, as it was of Lewis's [52], but Sosa [67] and Heller [34] propose only necessary conditions. Hence one may choose to read $K\varphi$ as "the agent *safely believes* φ/has *ruled out the relevant alternatives to* φ" for S/H-semantics. Our results for S/H-semantics can then be viewed as results about the logic of safe belief/the logic of relevant alternatives. However, for reasons similar to those given by Brueckner [10] and Murphy [57], if the subjunctivist or RA conditions are treated as necessary for knowledge, then closure failures for these conditions threaten closure for knowledge itself.[28] It is up to defenders of these theories to explain why knowledge is closed in ways that their conditions on knowledge are not.

[28] Suppose that C is a necessary but insufficient condition for knowledge, and let $C\varphi$ mean that the agent satisfies C with respect to φ. Hence $K\varphi \to C\varphi$ should be valid. Further suppose that (A) $C\varphi_1 \wedge \cdots \wedge C\varphi_n \to C\psi$ is not valid. As Vogel [73], Warfield [77], and others point out, it does not follow that (B) $K\varphi_1 \wedge \cdots \wedge K\varphi_n \to K\psi$ is not valid. For in the counterexample to (A), $K\varphi_1 \wedge \cdots \wedge K\varphi_n$ may not hold, since C is not sufficient for K.

Let C' be another insufficient condition such that C and C' are jointly sufficient for K, so $C\varphi \wedge C'\varphi \to K\varphi$ is valid. If (B) is valid, then $C'\varphi_1 \wedge \cdots \wedge C'\varphi_n$ does not hold in the counterexample to (A). Moreover, it must be that while (A) is not valid, $C\varphi_1 \wedge \cdots \wedge C\varphi_n \wedge C'\varphi_1 \wedge \cdots \wedge C'\varphi_n \to C\psi$ *is* valid. For if there is a counterexample to the latter, then there is a counterexample to (B), since C and C' are jointly sufficient and C is necessary for K.

The problem is that proposed conditions for K are typically independent in such a way that assuming one also satisfies C' with respect to $\varphi_1, \ldots, \varphi_n$ will not guarantee that one satisfies a distinct, non-redundant condition C with respect to ψ, if satisfying C with respect to $\varphi_1, \ldots, \varphi_n$ is not already sufficient. For example, if ruling out the relevant alternatives to $\varphi_1, \ldots, \varphi_n$ is not sufficient for ruling out the relevant alternatives to ψ, then what other condition is such that also satisfying it with respect to $\varphi_1, \ldots, \varphi_n$ will guarantee that one has ruled out the relevant alternatives to ψ? The same question arises for subjunctivist conditions. It is up to subjunctivists to say what they expect to block closure failures for knowledge, given closure failures for their necessary subjunctivist conditions on knowledge.

One way to do so is to build in the satisfaction of closure itself as another necessary condition. For example, Luper-Foy [53, 45n38] gives the "trivial example" of *contracking* φ, which is the condition (C') of satisfying the sensitivity condition (C) for all logical consequences of φ. However, this idea for building in closure misses the fact that multi-premise closure principles fail for contracking. For example, one can contrack p and contract q, while being *insensitive* with respect to $(p \wedge q) \vee r$ and therefore failing to contrack $p \wedge q$.

Contracking must be distinguished from another idea for combining tracking with closure. Roush [62, 63, Ch. 2, Section 1] proposes a disjunctive account according to which (to a first approximation) an agent knows ψ iff either the agent "Nozick-knows" ψ, i.e., satisfies Nozick's belief, sensitivity, and adherence conditions for ψ, or there are some $\varphi_1, \ldots, \varphi_n$ such that the agent knows $\varphi_1, \ldots, \varphi_n$ and knows that $\varphi_1 \wedge \cdots \wedge \varphi_n$ implies ψ (cf. Luper-Foy [53, 46] on "distracking"). Importantly, according to this *recursive tracking view of knowledge*, the tracking conditions (for which closure fails) are not necessary conditions for knowledge.

Definition 4.3 (**Truth in a CB Model**) Given a well-founded CB model $\mathcal{M} = \langle W, D, \leqslant, V \rangle$ with $w \in W$ and φ in the epistemic-doxastic language, define $\mathcal{M}, w \vDash_x \varphi$ as follows (with propositional cases as in Def. 3.4):

$$\begin{aligned}
&\mathcal{M}, w \vDash_x B\varphi \ \text{ iff } \forall v \in W\colon \text{if } wDv \text{ then } \mathcal{M}, v \vDash_x \varphi;\\
&\mathcal{M}, w \vDash_h K\varphi \ \text{ iff } \mathcal{M}, w \vDash_h B\varphi \text{ and}\\
&\qquad\qquad\qquad\qquad \text{(sensitivity) } \forall v \in \mathrm{Min}_{\leqslant_w}\left(\overline{[\![\varphi]\!]_h}\right) : \mathcal{M}, v \nvDash_h B\varphi;\\
&\mathcal{M}, w \vDash_n K\varphi \ \text{ iff } \mathcal{M}, w \vDash_n B\varphi \text{ and}\\
&\qquad\qquad\qquad\qquad \text{(sensitivity) } \forall v \in \mathrm{Min}_{\leqslant_w}\left(\overline{[\![\varphi]\!]_n}\right) : \mathcal{M}, v \nvDash_n B\varphi,\\
&\qquad\qquad\qquad\qquad \text{(adherence) } \forall v \in \mathrm{Min}_{\leqslant_w}\left([\![\varphi]\!]_n\right) : \mathcal{M}, v \vDash_n B\varphi;\\
&\mathcal{M}, w \vDash_s K\varphi \ \text{ iff } \mathcal{M}, w \vDash_s B\varphi \text{ and}\\
&\qquad\qquad\qquad\qquad \text{(safety) } \forall v \in \mathrm{Min}_{\leqslant_w}\left([\![B\varphi]\!]_s\right) : \mathcal{M}, v \vDash_s \varphi.
\end{aligned}$$

Note that the truth clause for $B\varphi$ guarantees *doxastic* closure (recall Section 2 and see Fact 5.11).[29]

It is easy to check that the belief and subjunctive conditions of H/N/S-semantics together ensure Fact 4.4 (cf. Heller [35, 126]; Kripke [44, 164]).

Fact 4.4 (**Veridicality**) *$K\varphi \to \varphi$ is H/N/S-valid.*

Observation 4.5 (**Adherence and Safety**) The adherence condition in the N-semantics clause may be equivalently replaced by

$$\forall v \in \mathrm{Min}_{\leqslant_w}(W)\colon \mathcal{M}, v \vDash_n \varphi \to B\varphi;$$

the safety condition in the S-semantics clause may be equivalently replaced by

$$\forall v \in \mathrm{Min}_{\leqslant_w}(W)\colon \mathcal{M}, v \vDash_s B\varphi \to \varphi.$$

This observation has two important consequences. The first is that in *centered* models (Def. 3.3.5), adherence ($\varphi \mathrel{\Box\!\!\!\to} B\varphi$) and safety ($B\varphi \mathrel{\Box\!\!\!\to} \varphi$) add nothing to belief and true belief, respectively, given standard Lewisian semantics for counterfactuals. DeRose [17, 27n27] takes adherence to be redundant apparently for this reason. But since we only assume *weak* centering, adherence as above makes a difference—obviously for truth in a model, but also for validity (see Fact 8.8). Nozick [58, 680n8]

[29] It is not essential here that we model belief with a doxastic accessibility relation. When we show that a given closure principle is H/N/S-*valid*, we use the fact that the truth clause for $B\varphi$ in Definition 4.3 guarantees some *doxastic closure* (see Fact 5.11); but when we show that a closure principle is *not* H/N/S-valid, we do not use any facts about doxastic closure, as one can verify by inspection of the proofs. For the purpose of demonstrating closure failures, we could simply associate with each $w \in W$ a set Σ_w of formulas such that $\mathcal{M}, w \vDash B\varphi$ iff $\varphi \in \Sigma_w$. However, if we were to assume no doxastic closure properties for Σ_w, then there would be no valid epistemic closure principles (except $K\varphi \to K\varphi$), assuming knowledge requires belief. As a modeling choice, this may be more realistic, but it throws away information about the reasons for closure failures. For we would no longer be able to tell whether an epistemic closure principle such as $K\varphi \to K(\varphi \vee \psi)$ is not valid for the (interesting) reason that the special conditions for knowledge posited by a theory are not preserved in the required way, or whether the principle is not valid for the (uninteresting) reason that there is some agent who knows φ but happened not to form a belief in $\varphi \vee \psi$.

suggests another way of understanding adherence so that it is non-trivial, but here I will settle on its simple interpretation with weak centering in standard semantics. Whether or not weak centering is right for counterfactuals, adherence and safety can be—and safety typically is—understood directly in terms of what holds in a set of close worlds including the actual world, our $\mathrm{Min}_{\leqslant_w}(W)$ (see note 26), rather than as $\varphi \mathrel{\Box\!\!\rightarrow} B\varphi$ and $B\varphi \mathrel{\Box\!\!\rightarrow} \varphi$.[30] (Adherence is often ignored). For sensitivity alone, centering vs. weak centering makes no difference for valid principles.

The second consequence is that safety is a $\exists\forall$ condition as in Section 3, where $\mathrm{Min}_{\leqslant_w}(W)$ serves as the set $\mathsf{R}_c(w)$ that is independent of the particular proposition in question (cf. Alspector-Kelly [3, 129n6]). By contrast, sensitivity is obviously a $\forall\exists$ condition, analogous to the D-semantics clause. Viewed this way, in the "subjunctivist-flavored" family of D/H/N/S-semantics, S-semantics is the odd member of the family, since by only looking at the fixed set $\mathrm{Min}_{\leqslant_w}(W)$ in the safety clause, it never uses the rest of the world-ordering.[31]

Figure 2 displays a CB model for Example 2.1. The dotted arrows represent the doxastic relation D. That the only arrow from w_1 goes to itself indicates that in w_1, student $\mathtt{A}$ believes that the actual world is w_1, where the patient has c and not x. (We do not require that D be functional, but in Fig. 2 it is.) Hence $\mathcal{M}, w_1 \vDash B(c \wedge \neg x)$. That the only arrow from w_3 goes to w_1 indicates that in w_3, $\mathtt{A}$ believes that w_1 is the actual world; since w_3 is the closest (to w_1) x-world, we take this to mean that if the patient's condition were x, $\mathtt{A}$ would still believe it was c and not x (because $\mathtt{A}$ did not run any of the tests necessary to detect x).[32] Hence $\mathcal{M}, w_1 \nvDash_{h,n} K\neg x$, because the *sensitivity* condition is violated. However, one can check that $\mathcal{M}, w_1 \vDash_{h,n} Kc$.

If we were to draw the model for student $\mathtt{B}$, we would replace the arrow from w_2 to w_2 by one from w_2 to w_1, reflecting that if the patient's condition were c', $\mathtt{B}$ would still believe it was c (because $\mathtt{B}$ made the diagnosis of c after only a physical exam,

[30] Alternatively, the sphere of worlds for adherence could be independent of the relation $\leqslant_w$ for sensitivity, i.e., distinct from $\mathrm{Min}_{\leqslant_w}(W)$ (see Holliday [38, Remark 3.2]), so $\leqslant_w$ could be centered without trivializing adherence. But this would allow cases in which an agent knows φ even though she believes φ in a $\neg\varphi$-world that is "close enough" to w to be in its adherence sphere (provided there is a closer $\neg\varphi$-world according to $\leqslant_w$ in which she does not believe φ). Nozick [58, 680n8] suggests interpreting adherence counterfactuals $\varphi \mathrel{\Box\!\!\rightarrow} B\varphi$ with true antecedents in such a way that the sphere over which $\varphi \rightarrow B\varphi$ must hold may differ for different φ. By contrast, Observation 4.5 shows that we are interpreting adherence as a kind of $\exists\forall$ condition, in a sense that generalizes that of Section 3 to cover a requirement that one meet an epistemic success condition in all P-worlds in $\mathsf{R}_c(w)$ (see Holliday [38, Section 3.3.2]). A $\forall\exists$ interpretation of adherence that, e.g., allows the adherence sphere for $\varphi \vee \psi$ to go beyond that of φ, would create another source of closure failure (see Sections 5.5 and 9).

[31] Note that safety and tracking theorists may draw different models, with different $\leqslant_w$ relations and $\mathrm{Min}_{\leqslant_w}(W)$ sets, to represent the epistemic situation of the same agent.

[32] What about w_4? In Section 3, we assumed that $\mathtt{A}$ learned from textbooks that x confers immunity to c, so she had eliminated w_4 at w_1. In Fig. 2, that the only arrow from w_4 goes to w_4 indicates that if (contrary to biological law) x did *not* confer immunity to c and the patient had both c and x, then $\mathtt{A}$ would believe that the patient had both c and x, perhaps because the textbooks and tests would be different in such a world. However, all we need to assume for the purposes of our example is that if the patient had both c and x, then it would be *compatible* with what $\mathtt{A}$ believes that the patient had both c and x, as indicated by the reflexive loop. We can have other outgoing arrows from w_4 as well.

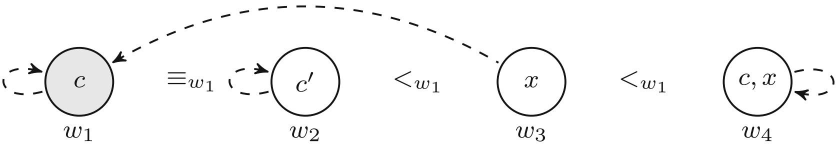

Fig. 2 A CB model for Example 2.1 (partially drawn)

and c and c' have the same visible symptoms). Hence $\mathcal{M}', w_1 \nvDash_{h,n} Kc$, where $\mathcal{M}'$ is the model with $w_2 D w_1$ instead of $w_2 D w_2$.

When we consider S-semantics, we get a different verdict on whether $\mathtt{A}$ knows that the patient does not have disease x. Observe that $\mathcal{M}, w_1 \vDash_s K\neg x$, because student $\mathtt{A}$ believes $\neg x$ in w_1 and at the closest worlds to w_1, namely w_1 and w_2, $\neg x$ is true. Therefore, $\mathtt{A}$ safely believes $\neg x$ in w_1. Similarly $\mathcal{M}, w_1 \vDash_s Kc$, because $\mathtt{A}$ safely believes c in w_1. Yet if we add the arrow from w_2 to w_1 for $\mathtt{B}$, one can check that $\mathtt{B}$ does not safely believe c at w_1, so $\mathcal{M}', w_1 \nvDash_s Kc$.

Remark 4.6 (*Vacuous Knowledge Again*) The fact that $\mathcal{M}, w_1 \vDash_s K\neg x$ reflects the idea that the safety theory leads to a *neo-Moorean* response to skepticism [67], according to which agents can know that skeptical hypotheses do not obtain. In general, a point parallel to that of Remark 3.9 holds for the $\mathsf{RS}_{\exists\forall}$ safety theory: if the $\neg\varphi$-worlds are not among the close worlds, then one's belief in φ is automatically safe, no matter how poorly one's beliefs match the facts in possible worlds (cf. Alspector-Kelly's [3] distinction between near-safe and far-safe beliefs). This is the version of the *problem of vacuous knowledge* for the safety theory (see Holliday [40]). By contrast, on the kind of $\mathsf{RS}_{\forall\exists}$ theory represented by H/N-semantics, if $\neg\varphi$ is possible, then knowledge requires that one not falsely believe φ in the closest $\neg\varphi$-worlds.

Like D-semantics, H/N-semantics avoid the skepticism of C-semantics and the vacuous knowledge of L/S-semantics, but at a cost for closure. All of the closure principles shown in Facts 3.10 and 3.11 to be falsifiable in RA models under D-semantics are also falsifiable in CB models under H/N-semantics, as one can check at w_1 in Fig. 2. After embracing the "nonclosure" of knowledge under known implication, Nozick [58, 231ff] tried to distinguish successful from unsuccessful cases of knowledge transmission by whether extra subjunctive conditions hold;[33] but doing so does not eliminate the unsuccessful cases, which go far beyond nonclosure under known implication, as shown in Section 5.

[33] Roughly, Nozick [58, 231ff] proposes than an agent knows ψ via inference from φ iff (1) $K\varphi$, (2) she infers the true conclusion ψ from premise φ, (3) $\neg\psi \;\Box\!\!\rightarrow \neg B\varphi$, and (4) $\psi \;\Box\!\!\rightarrow B\varphi$. Whether this proposal is consistent with the rest of Nozick's theory depends on whether (1)–(4) ensure that the agent tracks ψ, which is still necessary for her to know ψ (234); and that depends on what kind of modal connection between $B\varphi$ and $B\psi$ is supposed to follow from (2), because (1), (3), and (4) together do not ensure that she tracks ψ.

Nozick was well aware that $K(\varphi \wedge \psi) \rightarrow K\varphi \wedge K\psi$ fails on his theory, and his explanation (beginning "S's belief that *p&q* ..." on 228) is similar to a proof in our framework. He resisted the idea that $K\varphi \rightarrow K(\varphi \vee \psi)$ fails, but he is clearly committed to it.[34] Vogel's [76, 76] explanation of why it fails for Nozick is also similar to a proof in our framework, as are Kripke's [44] many demonstrations of closure failure for Nozick's theory. Partly in response to these problems, Roush [62, 63] proposes a *recursive tracking* view of knowledge, in a probabilistic framework, with an additional recursion clause to support closure (see note 26). For discussion of the relation between probabilistic and subjunctivist versions of tracking, see Holliday [38, Section 2.E].

All of the closure principles noted fail for S-semantics as well. For example, it is easy to construct a model in which $B(\varphi \wedge \psi)$ and hence $B\varphi$ are true at a world w, all worlds close to w satisfy $B(\varphi \wedge \psi) \rightarrow \varphi \wedge \psi$, and yet some worlds close to w do not satisfy $B\varphi \rightarrow \varphi$, resulting in a failure of $K(\varphi \wedge \psi) \rightarrow K\varphi$ at w. Murphy's [56, 57, Section 4.3] intuitive examples of closure failure for safety have exactly this structure.[35] We return to this problem for safety in Section 9.

Now it is time to go beyond case-by-case assessment of closure principles. In the following sections, we will turn to results of a more general nature.

5 The Closure Theorem and Its Consequences

In this section, I state the main result of the paper, Theorem 5.2, which characterizes the closure properties of knowledge for the theories we have formalized. Despite the differences between the RA, tracking, and safety theories of knowledge as formalized by D/H/N/S-semantics, Theorem 5.2 provides a unifying perspective: the valid epistemic closure principles are essentially the same for these different theories, except for a twist with the theory of *total* RA models. For comparison, I also include C/L-semantics, which fully support closure.

Formally, Theorem 5.2 is the same type of result as the "modal decomposition" results of van Benthem [8, Section 4.3, 10.4] for the weakest normal modal logic **K** and the weakest monotonic modal logic **M** (see Chellas [12, Section 8.2]). From Theorem 5.2 we obtain decidability (Corollary 5.9) and completeness (Corollary 7.1) results as corollaries. From the proof of the theorem, we obtain results on finite models (Corollary 5.24) and complexity (Corollary 5.25).

[34] While Nozick [58] admits that such a closure failure "surely carries things too far" (230n64, 692), he also says that an agent can know p and yet fail to know $\neg(\neg p \wedge \text{SK})$ (228). But the latter is logically equivalent to $p \vee \neg\, \text{SK}$, and Nozick accepts closure under (known) logical equivalence (229). Nozick suggests (236) that closure under deducing a disjunction from a disjunct should hold, provided methods of belief formation are taken into account. However, Holliday [38, Section 2.D] shows that taking methods into account does not help here.

[35] For Murphy's [57, Section 4.3] "Lying Larry" example, take φ to be *Larry is married* and ψ to be *Larry is married to Pat*. For Murphy's [56, 333] variation on Kripke's red barn example, take φ to be *the structure is a barn* and ψ to be *the structure is red*.

The following notation will be convenient throughout this section.

Notation 5.1 (**Closure Notation**) Given (possibly empty) sequences of formulas $\varphi_1, \dots, \varphi_n$ and $\psi_1, \dots, \psi_m$ in the epistemic language and a propositional conjunction φ_0, we use the notation

$$\chi_{n,m} := \varphi_0 \wedge K\varphi_1 \wedge \cdots \wedge K\varphi_n \to K\psi_1 \vee \cdots \vee K\psi_m.$$

Call such a $\chi_{n,m}$ a *closure principle*.[36]

Hence a closure principle states that if the agent knows each of φ_1 through φ_n (and the world satisfies a non-epistemic φ_0), then the agent knows at least one of ψ_1 through ψ_m. Our question is: which closure principles are *valid*?

Theorem 5.2 is the answer. Its statement refers to a "T-unpacked" closure principle, a notion not yet introduced. For the first reading of the theorem, think only of *flat* formulas $\chi_{n,m}$ without nesting of the K operator (Def. 2.2), which are T-unpacked if $\varphi_1 \wedge \cdots \wedge \varphi_n$ is a conjunct of φ_0. Or we can ignore T-unpacking for flat $\chi_{n,m}$ and replace condition (a) of the theorem by

(a)′ $\varphi_0 \wedge \cdots \wedge \varphi_n \to \bot$ is valid.

Example 5.8 will show the need for T-unpacking, defined in general in Section 5.2.1.

Theorem 5.2 (**Closure Theorem**) *Let*

$$\chi_{n,m} := \varphi_0 \wedge K\varphi_1 \wedge \cdots \wedge K\varphi_n \to K\psi_1 \vee \cdots \vee K\psi_m$$

be a T-unpacked closure principle.

1. $\chi_{n,m}$ *is C/L-valid over* relevant alternatives *models iff*

   (*a*) $\varphi_0 \to \bot$ *is valid or*
   (*b*) *for some* $\psi \in \{\psi_1, \dots, \psi_m\}$,

   $$\varphi_1 \wedge \cdots \wedge \varphi_n \to \psi \textit{ is valid};$$

2. $\chi_{n,m}$ *is D-valid over* total relevant alternatives *models iff* (*a*) *or*

   (*c*) *for some* $\Phi \subseteq \{\varphi_1, \dots, \varphi_n\}$ *and nonempty* $\Psi \subseteq \{\psi_1, \dots, \psi_m\}$,[37]

   $$\bigwedge_{\varphi \in \Phi} \varphi \leftrightarrow \bigwedge_{\psi \in \Psi} \psi \textit{ is valid};$$

3. $\chi_{n,m}$ *is D-valid over* all relevant alternatives *models iff* (*a*) *or*

   (*d*) *for some* $\Phi \subseteq \{\varphi_1, \dots, \varphi_n\}$ *and* $\psi \in \{\psi_1, \dots, \psi_m\}$,

   $$\bigwedge_{\varphi \in \Phi} \varphi \leftrightarrow \psi \textit{ is valid}.$$

[36] Following standard convention, we take an empty disjunction to be $\bot$, so a closure principle $\chi_{n,0}$ with no $K\psi$ formulas is of the form $\varphi_0 \wedge K\varphi_1 \wedge \cdots \wedge K\varphi_n \to \bot$.

[37] Following standard convention, if $\Phi = \emptyset$, we take $\bigwedge_{\varphi \in \Phi} \varphi$ to be $\top$.

4. $\chi_{n,m}$ *is H/N/S-valid over* counterfactual belief *models if* (a) *or* (d)*;*[38] *and a* flat $\chi_{n,m}$ *is H/N/S-valid over such models only if* (a) *or* (d)*.*

The remarkable fact established by Theorem 5.2 that D/H/N/S-semantics validate essentially the same closure principles, except for the twist of totality in (c), further supports talk of their representing a "family" of subjunctivist-flavored theories of knowledge. Although results in Section 8.2 (Facts 8.8.4, 8.8.5, and 8.10.1) show that the 'only if' direction of part 4 does not hold for some principles involving higher-order knowledge, the agreement between D/H/N/S-semantics on the validity of flat closure principles is striking.

Remark 5.3 (*Independence from Assumptions*) Recalling the types of orderings in Definition 3.3, it is noteworthy that parts 1 and 4 of Theorem 5.2 are independent of whether we assume totality (or universality), while parts 2 and 3 are independent of whether we assume centering, linearity (see Section 5.2.2), or universality (see Prop. 5.23). For parts 1–4, we can drop our running assumption of well-foundedness, provided we modify the truth definitions accordingly (see Remark 5.13). Finally, part 1 for L-semantics (but not C-semantics) and parts 2–3 for D-semantics are independent of additional properties of $\rightarrowtriangle$ such as transitivity and symmetry (see Remark 5.20 and Example 8.1).

To apply the theorem, observe that $Kp \wedge K(p \to q) \to Kq$ is not D/H/N/S-valid, because $p \wedge (p \to q) \to \bot$ is not valid, so (a)$'$ fails, and none of

$$p \wedge (p \to q) \leftrightarrow q,\;\; p \leftrightarrow q,\;\; (p \to q) \leftrightarrow q,\; \text{ or } \top \leftrightarrow q$$

are valid, so there are no Φ and Ψ/ψ as described. Hence (c)/(d) fails.

On the other hand, we now see that $Kp \wedge Kq \to K(p \wedge q)$ *is* D/H/N/S-valid, because $p \wedge q \leftrightarrow p \wedge q$ is valid, so we can take $\Phi = \{p, q\}$ and $\Psi = \{p \wedge q\}$ or $\psi = p \wedge q$. Besides $K\varphi \to \varphi$ (Facts 3.7 and 4.4), this is the first *valid* principle we have identified for D/H/N/S-semantics, to which we will return in Section 7.

Fact 5.4 (**C Axiom**) *The principle* $K\varphi \wedge K\psi \to K(\varphi \wedge \psi)$*, known as the C axiom, is D/H/N/S-valid.*

To get a feel for Theorem 5.2, it helps to test a variety of closure principles.

Exercise 5.5 (**Testing Closure**) Using Theorem 5.2, verify that neither $K(p \wedge q) \to K(p \vee q)$ nor $Kp \wedge Kq \to K(p \vee q)$ are D/H/N/S-valid; verify that $K(p \wedge q) \to Kp \vee Kq$ is only D-valid over *total* RA models; verify that $K(p \vee q) \wedge K(p \to q) \to Kq$ and $Kp \wedge K(p \to q) \to K(p \wedge q)$ are D/H/N/S-valid.

[38] When I refer to (d) from part 4, I mean the condition that $\bigwedge_{\varphi \in \Phi} \varphi \leftrightarrow \psi$ is H/N/S-valid.

As if the closure failures of Fact 3.11 were not bad enough, the first three of Exercise 5.5 are also highly counterintuitive. Recall from Section 2 that the Dretske-Nozick case against full closure under known implication, K, had two parts: examples in which K purportedly fails, such as Example 2.1, and theories of knowledge that purportedly explain the failures. For the other principles, we can see why they fail according to the subjunctivist-flavored theories; but without some intuitive examples in which, e.g., arguably an ideally astute logician knows two propositions but not their disjunction, the failure of such weak closure principles according to a theory of knowledge seems to be strong evidence against the theory—even for those sympathetic to the denial of K.

While the closure failures permitted by subjunctivist-flavored theories go too far, in another way they do not go far enough for some purposes. Reflection on the last two principles of Exercise 5.5 suggests they are about as dangerous as K in arguments for radical skepticism about knowledge. The fact that one's theory validates these principles seems to undermine the force of one's denying K in response to skepticism, as Nozick [58] uses his subjunctivism to do.

Notwithstanding these negative points against subjunctivist-flavored theories of *knowledge*, simply replace the K symbol in our language by a neutral $\Box$ and Theorem 5.2 can be viewed as a neutral result about the logic of relevant alternatives, of sensitive/truth-tracking belief, and of safe belief (see Section 7).

Parts 3 and 4 of Theorem 5.2 reflect that D-semantics over RA models and H/N/S-semantics over CB models have the following *separation property*.

Proposition 5.6 (**Separation**) *For D-semantics (resp. H/N/S-semantics), a closure principle $\chi_{n,m}$ (resp. a flat $\chi_{n,m}$) as in Notation 5.1 with $m \geq 1$ is valid iff there is some $j \leq m$ such that $\varphi_0 \wedge K\varphi_1 \wedge \cdots \wedge K\varphi_n \to K\psi_j$ is valid.*

The reason for this separation property comes out clearly in the proofs in Sections 5.3 and 5.4. In essence, if the principles with single disjunct consequents are all invalid, then we can glue their falsifying models together to obtain a falsifying model for $\chi_{n,m}$. However, this is not the case for D-semantics over *total* RA models. The following fact demonstrates the nonequivalence of D-semantics over total RA models and D-semantics over all RA models (as well as H/N/S-semantics over total/all CB models) with an interesting new axiom.

Fact 5.7 (**X Axiom**) *The principle $K(\varphi \wedge \psi) \to K\varphi \vee K\psi$, hereafter called the "X axiom" (see* Section 7*), is D-valid over* total *RA models, but not D-valid over all RA models or H/N/S-valid over (total) CB models.*

Proof I leave D-validity over total RA models to the reader. Figure 3 displays a non-total RA model that falsifies $K(p \wedge q) \to Kp \vee Kq$ in D-semantics. Since $\mathrm{Min}_{\preceq_w}\left(\overline{\llbracket p \wedge q \rrbracket}\right) = \{v, x\}$, $w \not\rightarrowtriangle v$, and $w \not\rightarrowtriangle x$, $\mathcal{M}, w \vDash_d K(p \wedge q)$. Since u and x are incomparable according to $\preceq_w$, as are y and v, we have $u \in \mathrm{Min}_{\preceq_w}\left(\overline{\llbracket p \rrbracket}\right)$ and $y \in \mathrm{Min}_{\preceq_w}\left(\overline{\llbracket q \rrbracket}\right)$, which with $w \rightarrowtriangle u$ and $w \rightarrowtriangle y$ implies $\mathcal{M}, w \nvDash_d Kp \vee Kq$. The counterexample for H/N/S-semantics is in Fig. 10, discussed in Section 9. □

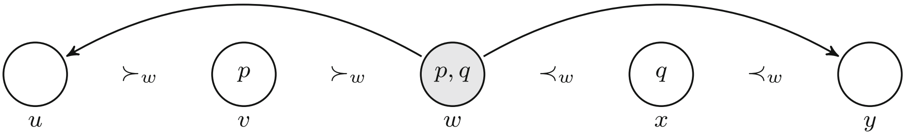

Fig. 3 A non-total RA countermodel for $K(p \wedge q) \to Kp \vee Kq$ in D-semantics (partially drawn, reflexive loops omitted)

In Section 7, we will see the role that the X axiom plays in a complete deductive system for D-semantics over total RA models, as well as the role that the C axiom plays in complete deductive systems for D/H/N/S-semantics.

Given the separation property, the proof of the 'only if' direction of Theorem 5.2.3 for *flat* closure principles can be explained roughly as follows.

Proof sketch Let us try to falsify a flat $\varphi_0 \wedge K\varphi_1 \wedge \cdots \wedge K\varphi_n \to K\psi_j$. Construct a pointed model $\mathcal{M}, w$ with a valuation such that the propositional part φ_0 is true at w.[39] To make $K\psi_j$ false while keeping all $K\varphi_i$ true at w, we want to add an *uneliminated* $\neg\psi_j$-world v such that (A) there is no $\neg\psi_j$-world more relevant than v and (B) for any $\neg\varphi_i$ true at v, there is a *more relevant* $\neg\varphi_i$-world that is *eliminated* at w. This is possible if there is a propositional valuation such that $\neg\psi_j$ is true at v and for all $\neg\varphi_i$ true at v, $\psi_j \wedge \neg\varphi_i$ is satisfiable; for then we can add a satisfying world for each conjunction and make them eliminated and more relevant than v, which gives (A) and (B). If there is no such valuation, then every valuation that satisfies $\neg\psi_j$ also satisfies some $\neg\varphi_i$ for which $\psi_j \to \varphi_i$ is valid. Then where Φ is the set of all such φ_i, $\neg\psi_j \to \bigvee_{\varphi\in\Phi} \neg\varphi$ and $\psi_j \to \bigwedge_{\varphi\in\Phi} \varphi$ are valid, which means $\bigwedge_{\varphi\in\Phi} \varphi \leftrightarrow \psi_j$ is valid. □

In Sections 5.2 and 5.3 we give a more precise and general form of the above argument. We conclude this subsection with an example of why Theorem 5.2 requires the notion of T-unpacking, which is defined in general in Definition 5.15.

Example 5.8 (*T-unpacking*) As noted before Theorem 5.2, if we consider only flat formulas, then we can ignore T-unpacking, provided we replace condition (a) of Theorem 5.2 by the condition: (a)$'$ $\varphi_0 \wedge \cdots \wedge \varphi_n \to \bot$ is valid. Let us see why T-unpacking is necessary for non-flat formulas. For example, the formula

$$KKp \wedge KKq \to K(p \wedge q) \tag{5.1}$$

is D/H/N/S-valid. Yet none of the following are valid: $Kp \wedge Kq \to \bot$, $Kp \wedge Kq \leftrightarrow p \wedge q$, $Kp \leftrightarrow p \wedge q$, $Kq \leftrightarrow p \wedge q$, and $\top \leftrightarrow p \wedge q$. Hence (5.1) does not satisfy (a)$'$,

[39] In the following argument, 'relevant' means relevant *at* w (i.e., according to $\preceq_w$) and 'uneliminated'/'eliminated' means uneliminated/eliminated *at* w (i.e., $w \rightarrowtriangle v$ or $w \not\rightarrowtriangle v$).

(c), or (d) in Theorem 5.2. However, if we *T-unpack* (5.1) by repeatedly applying the T axiom, $K\varphi \to \varphi$, to the antecedent, we obtain

$$(p \wedge q \wedge Kp \wedge Kq \wedge KKp \wedge KKq) \to K(p \wedge q), \tag{5.2}$$

which satisfies (b), (c), and (d) with $\Phi = \{p, q\}$ and $\Psi = \{p \wedge q\}$ or $\psi = p \wedge q$. Hence (5.2) is valid according to Theorem 5.2. Given the validity of the T axiom over RA/CB models (Facts 3.7 and 4.4), (5.1) and (5.2) are equivalent, so (5.1) is valid as well. This example shows the essential idea of T-unpacking, defined formally in Section 5.2.1 and demonstrated again in Example 5.17.

As shown by Proposition 5.16 below, any epistemic formula can be effectively transformed into an equivalent conjunction, each conjunct of which is a T-unpacked formula $\chi_{n,m}$ as in Notation 5.1. Using Theorem 5.2, the validity of each conjunct can be reduced to the validity of finitely many formulas of lesser modal depth (Def. 2.2). By repeating this process, we eventually obtain a finite set of propositional formulas, whose validity we can decide by truth tables. Thus, Theorem 5.2 yields the following decidability results.

Corollary 5.9 (Decidability) *The problem of checking whether an arbitrary formula is C/L/D-valid or whether a flat formula is H/N/S-valid over (total or all) RA/CB models is decidable.*

In addition, Theorem 5.2 will yield axiomatization results in Corollary 7.1. As Corollary 7.1 will show, the 'if' direction of each 'iff' statement in Theorem 5.2 is a soundness result, while the 'only if' direction is a completeness result. We prove soundness in Section 5.1 and completeness in Sections 5.2–5.4.

5.1 Soundness

In the 'if' direction, part 1 of Theorem 5.2 is a simple application of the C/L-truth definitions, which we skip. For parts 2–4, we use the following lemma.

Lemma 5.10 (Min Inclusion)

1. *If condition (c) of Theorem 5.2 holds, then for any well-founded and total pointed RA/CB model $\mathcal{M}, w$,*[40] *there is some $\psi \in \Psi$ such that*

$$\mathrm{Min}_{\leq_w}\left(\overline{[\![\psi]\!]}\right) \subseteq \bigcup_{\varphi \in \Phi} \mathrm{Min}_{\leq_w}\left(\overline{[\![\varphi]\!]}\right).$$

2. *If condition (d) of Theorem 5.2 holds, then for any well-founded pointed RA/CB model $\mathcal{M}, w$,*

$$\mathrm{Min}_{\leq_w}\left(\overline{[\![\psi]\!]}\right) \subseteq \bigcup_{\varphi \in \Phi} \mathrm{Min}_{\leq_w}\left(\overline{[\![\varphi]\!]}\right).$$

[40] When dealing with both RA and CB models, I use $\leq_w$ to stand for $\preceq_w$ or $\leqslant_w$.

Proof For part 1, assume for reductio that (c) holds and there is some well-founded and total $\mathcal{M}, w$ such that for all $\psi \in \Psi$ there is some u_ψ with

$$u_\psi \in \operatorname{Min}_{\leq_w}\left(\overline{[\![\psi]\!]}\right) \tag{5.3}$$

and

$$u_\psi \notin \bigcup_{\varphi\in\Phi} \operatorname{Min}_{\leq_w}\left(\overline{[\![\varphi]\!]}\right). \tag{5.4}$$

Given (c), (5.3) implies $u_\psi \in \overline{[\![\varphi_\psi]\!]}$ for some $\varphi_\psi \in \Phi$. Since $\leq_w$ is well-founded, there is some

$$v \in \operatorname{Min}_{\leq_w}\left(\bigcup_{\varphi\in\Phi}\overline{[\![\varphi]\!]}\right). \tag{5.5}$$

Given (c), (5.5) implies $v \in \overline{[\![\psi]\!]}$ for some $\psi \in \Psi$. Hence $u_\psi \leq_w v$ by (5.3) and the totality of $\leq_w$. Together $u_\psi \leq_w v$, $u_\psi \in \overline{[\![\varphi_\psi]\!]}$, (5.5), and the transitivity of $\leq_w$ imply

$$u_\psi \in \operatorname{Min}_{\leq_w}\left(\bigcup_{\varphi\in\Phi}\overline{[\![\varphi]\!]}\right), \tag{5.6}$$

which contradicts (5.4) by basic set theory.

For part 2, assume for reductio that (d) holds and there is some well-founded $\mathcal{M}, w$ and u_ψ such that (5.3) and (5.4) hold for ψ. Given (d), (5.3) implies $u_\psi \in \overline{[\![\varphi_\psi]\!]}$ for some $\varphi_\psi \in \Phi$. Hence by the well-foundedness of $\leq_w$ and (5.4) there is some $v \in \overline{[\![\varphi_\psi]\!]}$ such that $v <_w u_\psi$. Given (d), $v \in \overline{[\![\varphi_\psi]\!]}$ implies $v \in \overline{[\![\psi]\!]}$, which with $v <_w u_\psi$ contradicts (5.3). □

For the H/N/S-semantics cases, we will also use a basic fact of normal modal logic (see Theorem 3.3(2) of Chellas [12]), namely that the truth clause for B in Definition 4.3 guarantees Fact 5.11 below. Note that we do not require full doxastic closure, but only as much doxastic closure as needed to support the limited forms of epistemic closure that are valid for H/N/S-semantics.

Fact 5.11 (Partial Doxastic Closure) *For $x \in \{h, n, s\}$, if $\vDash_x \bigwedge_{\varphi\in\Phi} \varphi \leftrightarrow \psi$, then $\vDash_x \bigwedge_{\varphi\in\Phi} B\varphi \leftrightarrow B\psi$.*

For convenience, we will use the following notation throughout this section.

Notation 5.12 (Relational Image) Given $\mathcal{M} = \langle W, \rightarrowtriangle, \preceq, V\rangle$, the image of $\{w\}$ under the relation $\rightarrowtriangle$ is $\rightarrowtriangle(w) = \{v \in W \mid w \rightarrowtriangle v\}$.

Hence $\rightarrowtriangle(w)$ is the set of uneliminated possibilities for the agent in w.

We are now ready to prove the 'if' directions of Theorem 5.2.2–4.

Claim If (a) or (c) holds, then $\chi_{n,m}$ is D-valid over total RA models; if (a) or (d) holds, then it is D-valid over RA models and H/N/S-valid over CB models.

Proof If (a) holds, then it is immediate that $\chi_{n,m}$ is D/H/N/S-valid, since its antecedent is always false. For (c) and (d), we consider each of the D/H/N/S-semantics in turn, assuming for an arbitrary pointed RA/CB model $\mathcal{M}, w$ that

$$\mathcal{M}, w \vDash_x \bigwedge_{\varphi \in \Phi} K\varphi. \tag{5.7}$$

To show $\mathcal{M}, w \vDash_x \chi_{n,m}$, it suffices to show $\mathcal{M}, w \vDash_x K\psi_j$ for some $j \leq m$.

If (5.7) holds for $x := d$, then by the truth definition (Def. 3.6),

$$\bigcup_{\varphi \in \Phi} \mathrm{Min}_{\preceq_w}\left(\overline{[\![\varphi]\!]}\right) \cap \rightarrowtriangle(w) = \emptyset. \tag{5.8}$$

If $\mathcal{M}$ is a total (resp. any) RA model, then by (c) and Lemma 5.10.1 (resp. by (d) and Lemma 5.10.2), (5.8) implies that there is some $\psi \in \Psi$ (resp. that the ψ in (d) is) such that $\mathrm{Min}_{\leq_w}(\overline{[\![\psi]\!]}) \cap \rightarrowtriangle(w) = \emptyset$ whence $\mathcal{M}, w \vDash_d K\psi$.

For the cases of H/N/S-semantics, it follows from (d) and Fact 5.11 that

$$\bigcap_{\varphi \in \Phi} [\![B\varphi]\!] = [\![B\psi]\!] \text{ and } \bigcup_{\varphi \in \Phi} \overline{[\![B\varphi]\!]} = \overline{[\![B\psi]\!]}. \tag{5.9}$$

If (5.7) holds for $x := h$, then by the truth definition (Def. 4.3),

$$\mathcal{M}, w \vDash_h \bigwedge_{\varphi \in \Phi} B\varphi \text{ and } \bigcup_{\varphi \in \Phi} \mathrm{Min}_{\leqslant_w}\left(\overline{[\![\varphi]\!]}\right) \subseteq \bigcup_{\varphi \in \Phi} \overline{[\![B\varphi]\!]}. \tag{5.10}$$

By (5.9), the first conjunct of (5.10) implies $\mathcal{M}, w \vDash_h B\psi$. By (d), Lemma 5.10.2, and (5.9), the second conjunct implies the sensitivity condition that $\mathrm{Min}_{\leqslant_w}\left(\overline{[\![\psi]\!]}\right) \subseteq \overline{[\![B\psi]\!]}$. Hence $\mathcal{M}, w \vDash_h K\psi$.

If (5.7) holds for $x := n$, then by the truth definition (Def. 4.3), (5.10) holds with n in place of h. So by the same argument as before, sensitivity holds for ψ at w, which with $\mathcal{M}, w \vDash_n B\psi$ and $w \in \mathrm{Min}_{\leqslant_w}(W)$ (Def. 3.1.3b) implies $\mathcal{M}, w \vDash_n \psi$. It follows that $\mathrm{Min}_{\leqslant_w}([\![\psi]\!]) \subseteq \mathrm{Min}_{\leqslant_w}(W)$, which with (d) implies

$$\mathrm{Min}_{\leqslant_w}([\![\psi]\!]) \subseteq \bigcap_{\varphi \in \Phi} \mathrm{Min}_{\leqslant_w}([\![\varphi]\!]). \tag{5.11}$$

Since the adherence condition must hold for each $\varphi \in \Phi$ at w,

$$\bigcap_{\varphi \in \Phi} \mathrm{Min}_{\leqslant_w}([\![\varphi]\!]) \subseteq \bigcap_{\varphi \in \Phi} [\![B\varphi]\!], \tag{5.12}$$

which with (5.11) and (5.9) implies $\mathrm{Min}_{\leqslant_w}([\![\psi]\!]) \subseteq [\![B\psi]\!]$. Thus, adherence and sensitivity hold for ψ at w, so $\mathcal{M}, w \vDash_n K\psi$ given $\mathcal{M}, w \vDash_n B\psi$.

If (5.7) holds for $x := s$, then by the truth definition (Def. 4.3),

$$\mathcal{M}, w \vDash_s \bigwedge_{\varphi \in \Phi} B\varphi \text{ and } \bigcap_{\varphi \in \Phi} \mathrm{Min}_{\leqslant_w}([\![B\varphi]\!]) \subseteq \bigcap_{\varphi \in \Phi} [\![\varphi]\!]. \tag{5.13}$$

By (5.9), the first conjunct of (5.13) implies $\mathcal{M}, w \vDash_s B\psi$. Given $w \in \text{Min}_{\leqslant_w}(W)$ (Def. 3.1.3b), it follows that $\text{Min}_{\leqslant_w}([\![B\psi]\!]) \subseteq \text{Min}_{\leqslant_w}(W)$ and therefore

$$\text{Min}_{\leqslant_w}([\![B\psi]\!]) \subseteq \bigcap_{\varphi\in\Phi} \text{Min}_{\leqslant_w}([\![B\varphi]\!]) \tag{5.14}$$

by (5.9). Finally, from (d) we have

$$\bigcap_{\varphi\in\Phi} [\![\varphi]\!] \subseteq [\![\psi]\!], \tag{5.15}$$

which with (5.14) and the second conjunct of (5.13) implies the safety condition that $\text{Min}_{\leqslant_w}([\![B\psi]\!]) \subseteq [\![\psi]\!]$, so $\mathcal{M}, w \vDash_s K\psi$ given $\mathcal{M}, w \vDash_s B\psi$. □

Remark 5.13 (*Dropping Well-Foundedness*) We can drop the assumption of well-foundedness used in the above proofs, provided we modify the truth definitions accordingly. For example (cf. Lewis [49, Section 2.3]), we may define

$$\mathcal{M}, w \vDash_{d'} K\varphi \text{ iff } \begin{cases} [\![\varphi]\!]_{d'} = W_w \text{ or} \\ \exists v \in \overline{[\![\varphi]\!]}_{d'} \cap W_w\ \forall u \in \overline{[\![\varphi]\!]}_{d'}\colon \text{if } u \preceq_w v \text{ then } w \not\rightarrowtriangle u, \end{cases} \tag{5.16}$$

which is equivalent to the clause in Definition 3.6 over (total) well-founded models.[41] I will give the proof for Theorem 5.2.2 that (c) implies the validity of $\chi_{n,m}$ over total RA models according to (5.16). Assume that (5.7) holds for $x := d'$. If $[\![\varphi]\!] = W_w$ for all $\varphi \in \Phi$, then by (c), $[\![\psi]\!] = W_w$ and hence $\mathcal{M}, w \vDash_{d'} K\psi$ for all $\psi \in \Psi$. Otherwise, for every $\varphi \in \Phi$ for which the second case of (5.16) holds, let v_φ be a witness to the existential quantifier. Since $\{v_\varphi \mid \varphi \in \Phi\}$ is finite and nonempty, $\text{Min}_{\preceq_w}(\{v_\varphi \mid \varphi \in \Phi\})$ is nonempty. Consider some $v \in \text{Min}_{\preceq_w}(\{v_\varphi \mid \varphi \in \Phi\})$. Given that $\preceq_w$ is a total preorder,

$$\forall u \in \bigcup_{\varphi\in\Phi} \overline{[\![\varphi]\!]}_{d'}\colon \text{if } u \preceq_w v \text{ then } w \not\rightarrowtriangle u. \tag{5.17}$$

Since $v \in \overline{[\![\varphi]\!]}$ for some $\varphi \in \Phi$, by (c) it follows that $v \in \overline{[\![\psi]\!]}$ for some $\psi \in \Psi$. Now observe that for all $u \in \overline{[\![\psi]\!]}$, $u \preceq_w v$ implies $w \not\rightarrowtriangle u$. For if $u \in \overline{[\![\psi]\!]}$, then by (c), $u \in \overline{[\![\varphi]\!]}$ for some $\varphi \in \Phi$, in which case $u \preceq_w v$ implies $w \not\rightarrowtriangle u$ by (5.17). Hence v is a witness to the existential in (5.16) for $K\psi$, whence $\mathcal{M}, w \vDash_{d'} K\psi$.

We leave the other cases without well-foundedness to the reader.[42]

[41] Equation (5.16) assumes totality. Without totality, we replace the right side of (5.16) with:

$$\forall x \in W_w\colon \text{if } x \in \overline{[\![\varphi]\!]} \text{ then } \exists v \in \overline{[\![\varphi]\!]}, v \preceq_w x, \forall u \in \overline{[\![\varphi]\!]}\colon \text{if } u \preceq_w v \text{ then } w \not\rightarrowtriangle u.$$

For the proof that (d) in Theorem 5.2.3 implies the validity of $\chi_{n,m}$ over all RA models according to this modified truth clause, see Holliday [38, Section 2.6.1].

[42] For H-semantics without well-foundedness (but with totality), define a new $\vDash_{h'}$ relation as in (5.16) but with $\mathcal{M}, u \nvDash_{h'} B\varphi$ in place of $w \not\rightarrowtriangle u$ and with the belief condition for knowledge. Then the proof of the 'if' direction of Theorem 5.2.4 for $\vDash_{h'}$ is similar to the proof above for $\vDash_{d'}$, but replacing (c) by (d) and replacing $w \not\rightarrowtriangle u$ in (5.17) by $\mathcal{M}, u \nvDash_{h'} B\psi$, which follows from $\mathcal{M}, u \nvDash_{h'} B\varphi$ for any $\varphi \in \Phi$ by (d) and Fact 5.11. Without totality, we use the truth clause for K from the previous footnote but with $\mathcal{M}, u \nvDash B\varphi$ in place of $w \not\rightarrowtriangle u$ and with the belief condition (see Holliday [38, Section 2.6.1]). Finally, since Definition 3.1.3b implies that $\text{Min}_{\leqslant_w}(W) \neq \emptyset$ even if $\leqslant_w$ is not well-founded, it follows from Observation 4.5 that the adherence and safety conditions of N/S-semantics do not require well-foundedness.

5.2 Completeness for Total RA Models

We turn now to the 'only if' directions of Theorem 5.2. The proof for part 1 of the theorem, which we omit, is a much simpler application of the general approach used for the other parts. In this section, we treat the 'only if' direction of part 2. This is the most involved part of the proof and takes us most of the way toward the 'only if' direction of part 3, treated in Section 5.3. It may help at times to recall the proof sketch given after Fact 5.7 above.

In Section 5.2.1, I define what it is for the $\chi_{n,m}$ in Theorem 5.2 to be *T-unpacked*. In Section 5.2.2, I show that if a T-unpacked $\chi_{n,m}$ does not satisfy (a) or (c) of Theorem 5.2, then it is falsified by a finite *total* RA model according to D-semantics. In fact, it is falsified by a finite *linear* RA model with the *universal field* property (Def. 3.3.4). Finally, in Section 5.2.3 we give upper bounds on the size of and complexity of finding falsifying models in Corollaries 5.24 and 5.25.

5.2.1 T-unpacking Formulas

Toward defining what it is for $\chi_{n,m}$ (Notation 5.1) to be T-unpacked, let us first define a normal form for the $\varphi_1, \ldots, \varphi_n$ in $\chi_{n,m}$. For our purposes, we need only define the normal form for the top (propositional) level of each φ_i.

Definition 5.14 (**DNF**) A formula in the epistemic language is in (propositional) *disjunctive normal form* (DNF) iff it is of the form

$$\bigvee \left(\alpha \wedge \bigwedge K\beta \wedge \bigwedge \neg K\gamma\right),$$

where α is propositional (a conjunction of literals, but it will not matter here), and β and γ are any formulas.

Roughly speaking, we T-unpack a conditional $\chi_{n,m}$ by using the T axiom, $K\varphi_i \to \varphi_i$, to replace $K\varphi_i$ in the antecedent with the equivalent $\varphi_i \wedge K\varphi_i$ and then use propositional logic to put φ_i in its appropriate place; e.g., if φ_i is $\neg K\gamma$, then we move $K\gamma$ to the consequent to become one of the $K\psi$'s. After the following general definition and result, we work out a concrete example.

Definition 5.15 (**T-unpacked**) For any (possibly empty) sequence of formulas $\psi_1, \ldots, \psi_m$, a formula of the form $\chi_{0,m}$ is T-unpacked; and for φ_{n+1} in DNF, a formula of the form $\chi_{n+1,m}$ is T-unpacked iff $\chi_{n,m}$ is T-unpacked and there is a disjunct δ of φ_{n+1} such that:

1. the α conjunct in δ is a conjunct of φ_0;
2. for all $K\beta$ conjuncts in δ, there is some $i \leq n$ such that $\varphi_i = \beta$;
3. for all $\neg K\gamma$ conjuncts in δ, there is some $j \leq m$ such that $\psi_j = \gamma$.

The following proposition will be used to prove several later results.

Proposition 5.16 (**T-unpacking**) *Every formula in the epistemic language is equivalent over RA models in C/D/L-semantics (and over CB models in H/N/S-semantics) to a conjunction of T-unpacked formulas of the form* $\chi_{n,m}$.

Proof By propositional logic, every formula θ is equivalent to a conjunction of formulas of the conditional (disjunctive) form $\chi_{n,m}$. Also by propositional logic, every φ_i in the antecedent of $\chi_{n,m}$ can be converted into an equivalent $\varphi_i^{\vee}$ in DNF; and since φ_i and $\varphi_i^{\vee}$ are equivalent, so are $K\varphi_i$ and $K\varphi_i^{\vee}$ by the semantics. To obtain an equivalent of θ in which each $\chi_{n,m}$ is T-unpacked, we repeatedly use the following equivalences, easily derived using propositional logic and the valid T axiom, $K\psi \to \psi$. Where ζ and η are any formulas,

$$\begin{aligned}
&\zeta \wedge K\left(\bigvee_{k\leq l}\delta_k\right) \to \eta\\
\Leftrightarrow\ &\zeta \wedge \left(\bigvee_{k\leq l}\delta_k\right) \wedge K\left(\bigvee_{k\leq l}\delta_k\right) \to \eta\\
\Leftrightarrow\ &\bigwedge_{k\leq l}\left(\zeta \wedge \delta_k \wedge K\left(\bigvee_{k\leq l}\delta_k\right) \to \eta\right)\\
\Leftrightarrow\ &\bigwedge_{k\leq l}\left(\zeta \wedge \alpha^k \wedge \bigwedge K\beta^k \wedge K\left(\bigvee_{k\leq l}\delta_k\right) \to \eta \vee \bigvee K\gamma^k\right),
\end{aligned}$$

where each δ_k is of the form $\alpha^k \wedge \bigwedge K\beta^k \wedge \bigwedge \neg K\gamma^k$. Compare conditions 1–3 of Definition 5.15 to the relation of δ_k to the k-th conjunct in the last line. □

Example 5.17 (*T-unpacking cont.*) Let us T-unpack the following formula:

$$\underline{K\left(\underline{\left(K\underline{(Kp\vee q)}_{\beta_1^1} \wedge K\underline{\neg Kq}_{\beta_2^1} \wedge \neg K\underline{Kr}_{\gamma_1^1}\right)}_{\delta_1} \vee \underline{K\underline{\neg Kr}_{\beta_1^2}}_{\delta_2}\right)}_{\varphi} \to K\psi.$$

No matter what we substitute for ψ, the form of the final result will be the same, since T-unpacking does nothing to formulas already in the consequent.

As in the proof of Proposition 5.16, we derive a string of equivalences, obtaining formulas in boldface by applications of the T axiom and otherwise using only propositional logic:

$$K\varphi \to K\psi \Leftrightarrow \boldsymbol{\varphi} \wedge K\varphi \to K\psi;$$

then since φ is a disjunction, we split into two conjuncts:

$$\Leftrightarrow \quad (\delta_1 \wedge K\varphi \to K\psi) \wedge (\delta_2 \wedge K\varphi \to K\psi);$$

then we move the negated $K\gamma_1^1$ in δ_1 to the first consequent and rewrite as

$$\Leftrightarrow \quad \left(K\beta_1^1 \wedge K\beta_2^1 \wedge K\varphi \to K\psi \vee K\gamma_1^1\right) \wedge \left(K\beta_1^2 \wedge K\varphi \to K\psi\right);$$

then we apply the T axiom to the $K\beta$ formulas:

$$\Leftrightarrow \quad \big(\boldsymbol{\beta_1^1} \wedge \boldsymbol{\beta_2^1} \wedge K\beta_1^1 \wedge K\beta_2^1 \wedge K\varphi \to K\psi \vee K\gamma_1^1\big) \wedge \big(\boldsymbol{\beta_1^2} \wedge K\beta_1^2 \wedge K\varphi \to K\psi\big);$$

then we move the negated Kq in β_2^1 and Kr in β_1^2 to the consequents:

$$\Leftrightarrow \quad \big(\beta_1^1 \wedge K\beta_1^1 \wedge K\beta_2^1 \wedge K\varphi \to K\psi \vee K\gamma_1^1 \vee Kq\big) \wedge (K\beta_1^2 \wedge K\varphi \to K\psi \vee Kr);$$

since β_1^1 is another disjunction, we split the first conjunct into two:

$$\begin{aligned} \Leftrightarrow \quad & \big(Kp \wedge K\beta_1^1 \wedge K\beta_2^1 \wedge K\varphi \to K\psi \vee K\gamma_1^1 \vee Kq\big) \wedge \\ & \big(q \wedge K\beta_1^1 \wedge K\beta_2^1 \wedge K\varphi \to K\psi \vee K\gamma_1^1 \vee Kq\big) \wedge \\ & \big(K\beta_1^2 \wedge K\varphi \to K\psi \vee Kr\big); \end{aligned}$$

finally, we apply the T axiom to Kp and rewrite as:

$$\begin{aligned} \Leftrightarrow \quad & \Big(\underline{\boldsymbol{p}}_{\varphi_0} \wedge K\underline{p}_{\varphi_1} \wedge K\underline{(Kp \vee q)}_{\varphi_2} \wedge K\underline{\neg Kq}_{\varphi_3} \wedge K\underline{\varphi}_{\varphi_4} \to K\underline{\psi}_{\psi_1} \vee K\underline{Kr}_{\psi_2} \vee K\underline{q}_{\psi_3}\Big) \\ & \wedge \Big(\underline{q}_{\varphi_0'} \wedge K\underline{(Kp \vee q)}_{\varphi_1'} \wedge K\underline{\neg Kq}_{\varphi_2'} \wedge K\underline{\varphi}_{\varphi_3'} \to K\underline{\psi}_{\psi_1'} \vee K\underline{Kr}_{\psi_2'} \vee K\underline{q}_{\psi_3'}\Big) \\ & \wedge \Big(K\underline{\neg Kr}_{\varphi_1''} \wedge K\underline{\varphi}_{\varphi_2''} \to K\underline{\psi}_{\psi_1''} \vee K\underline{r}_{\psi_2''}\Big). \end{aligned}$$

Observe that the three conjuncts are T-unpacked according to Definition 5.15.

5.2.2 *Countermodel Construction*

Our approach to proving the 'only if' direction of Theorem 5.2.2 is to assume that (a) and (c) fail, from which we infer the existence of models that can be "glued together" to construct a countermodel for $\chi_{n,m}$. For a clear illustration of this approach applied to basic modal models with arbitrary accessibility relations, see van Benthem [8, Section 4.3]. There are two important differences in what we must do here. First, since we are dealing with reflexive models in which $K\varphi \to \varphi$ is valid, we must use T-unpacking. Second, since we are dealing with a hybrid of relational and *ordering* semantics, we cannot simply glue all of the relevant models together at once, as in the basic modal case; instead, we must put them in the right order, which we do inductively.

The construction has two main parts. First, we inductively build up a kind of "pre-model" that falsifies $\chi_{n,m}$. Second, assuming that $\chi_{n,m}$ is T-unpacked, we can then convert the pre-model into an RA model that falsifies $\chi_{n,m}$.

Definition 5.18 (Pre-model) A pointed *pre-model* is a pair $\mathcal{M}, v$, with $\mathcal{M} = \langle W, \twoheadrightarrow, \preceq, V\rangle$ and $v \in W$, where W, $\twoheadrightarrow$, $\preceq_w$ for $w \in W \setminus \{v\}$, and V are as in Definition 3.1; $\preceq_v$ satisfies Definition 3.1.3a, but for all $w \in W$, $v \notin W_w$.

Hence a pointed pre-model is not a pointed RA model, since Definition 3.1.3b requires that $v \in W_v$ for an RA model. However, truth at a pointed pre-model is defined in the same way as truth at a pointed RA model in Definition 3.6.

The following lemma shows how we will build up our model in the inductive construction of Lemma 5.21. It is important to note that Lemmas 5.19 and 5.21 hold for any $\chi_{n,m}$ as in Notation 5.1, whether or not it is T-unpacked.

Lemma 5.19 (**Pre-model Extension**) *Assume there is a linear pointed pre-model* $\mathcal{M}, w$ *such that* $\mathcal{M}, w \nvDash_d \chi_{n,m}$.

1. *If* $\psi_1 \wedge \cdots \wedge \psi_m \to \varphi_{n+1}$ *is not D-valid over linear RA models, then there is a linear pointed pre-model* $\mathcal{M}^\sharp, w$ *such that* $\mathcal{M}^\sharp, w \nvDash_d \chi_{n+1,m}$.

2. *If* $\varphi_1 \wedge \cdots \wedge \varphi_n \to \psi_{m+1}$ *is not D-valid over linear RA models, then there is a linear pointed pre-model* $\mathcal{M}^\flat, w$ *such that* $\mathcal{M}^\flat, w \nvDash_d \chi_{n,m+1}$.

Proof For part 1, let $\mathcal{N} = \langle N, \rightarrowtriangle^{\mathcal{N}}, \preceq^{\mathcal{N}}, V^{\mathcal{N}} \rangle$ with $v \in N$ be a linear RA model such that $\mathcal{N}, v \nvDash_d \psi_1 \wedge \cdots \wedge \psi_m \to \varphi_{n+1}$. By assumption, there is a linear pre-model $\mathcal{M} = \langle M, \rightarrowtriangle^{\mathcal{M}}, \preceq^{\mathcal{M}}, V^{\mathcal{M}} \rangle$ with point $w \in M$ such that $\mathcal{M}, w \nvDash_d \chi_{n,m}$. Define $\mathcal{M}^\sharp = \langle W^\sharp, \rightarrowtriangle^\sharp, \preceq^\sharp, V^\sharp \rangle$ as follows (see Fig. 4):

$W^\sharp = M \cup N$ (we can assume $M \cap N = \emptyset$); $\rightarrowtriangle^\sharp = \rightarrowtriangle^{\mathcal{M}} \cup \rightarrowtriangle^{\mathcal{N}}$;
$\preceq^\sharp_w = \preceq^{\mathcal{M}}_w \cup \{\langle v, x \rangle \mid x = v \text{ or } x \in M_w\}$, where M_w is the field of $\preceq^{\mathcal{M}}_w$;
$\preceq^\sharp_x = \preceq^{\mathcal{M}}_x$ for all $x \in M \setminus \{w\}$; $\preceq^\sharp_y = \preceq^{\mathcal{N}}_y$ for all $y \in N$;
$V^\sharp(p) = V^{\mathcal{M}}(p) \cup V^{\mathcal{N}}(p)$.

Observe that $\mathcal{M}^\sharp, w$ is a linear pointed pre-model.

It is easy to verify that for all formulas ξ and $x \in M \setminus \{w\}$,

$$\mathcal{M}^\sharp, x \vDash_d \xi \text{ iff } \mathcal{M}, x \vDash_d \xi \text{ ; and } \mathcal{M}^\sharp, v \vDash_d \xi \text{ iff } \mathcal{N}, v \vDash_d \xi. \tag{5.18}$$

Given $\mathcal{M}, w \nvDash_d \chi_{n,m}$ and the truth definition (Def. 3.6),

$$\bigcup_{1 \le i \le n} \mathrm{Min}_{\preceq^{\mathcal{M}}_w} \left(\overline{[\![\varphi_i]\!]}^{\mathcal{M}} \right) \cap \rightarrowtriangle^{\mathcal{M}}(w) = \emptyset. \tag{5.19}$$

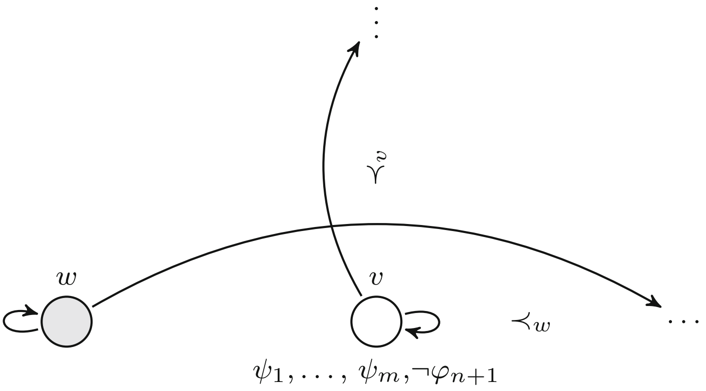

Fig. 4 Part of the extended pre-model $\mathcal{M}^\sharp$ for Lemma 5.19.1

It follows by the construction of $\mathcal{M}^\sharp$ and (5.18) that

$$\bigcup_{1\leq i\leq n+1} \mathrm{Min}_{\preceq^\sharp_w}\left(\overline{[\![\varphi_i]\!]}^{\mathcal{M}^\sharp}\right) \cap \rightarrowtriangle^\sharp(w) = \emptyset, \tag{5.20}$$

which is equivalent to $\mathcal{M}^\sharp, w \vDash_d K\varphi_1 \wedge \cdots \wedge K\varphi_{n+1}$ by the truth definition. The construction of $\mathcal{M}^\sharp$ and (5.18) also guarantee that for all $k \leq m$,

$$\mathrm{Min}_{\preceq^{\mathcal{M}}_w}\left(\overline{[\![\psi_k]\!]}^{\mathcal{M}}\right) \cap \rightarrowtriangle^{\mathcal{M}}(w) \subseteq \mathrm{Min}_{\preceq^\sharp_w}\left(\overline{[\![\psi_k]\!]}^{\mathcal{M}^\sharp}\right) \cap \rightarrowtriangle^\sharp(w). \tag{5.21}$$

Given $\mathcal{M}, w \nvDash_d \chi_{n,m}$, for all $k \leq m$ the left side of (5.21) is nonempty, so the right side is nonempty. Hence by the truth definition, $\mathcal{M}^\sharp, w \nvDash_d K\psi_1 \vee \cdots \vee K\varphi_m$. Finally, since φ_0 is propositional, $\mathcal{M}, w \vDash \varphi_0$ implies $\mathcal{M}^\sharp, w \vDash \varphi_0$ by definition of $V^\sharp$. It follows from the preceding facts that $\mathcal{M}^\sharp, w \nvDash_d \chi_{n+1,m}$.

For part 2, let $\mathcal{O} = \langle O, \rightarrowtriangle^{\mathcal{O}}, \preceq^{\mathcal{O}}, V^{\mathcal{O}}\rangle$ with $u \in O$ be a linear RA model such that $\mathcal{O}, u \nvDash_d \varphi_1 \wedge \cdots \wedge \varphi_n \rightarrow \psi_{m+1}$. Given $\mathcal{M}, w$ as in part 1, define $\mathcal{M}^\flat = \langle W^\flat, \rightarrowtriangle^\flat, \preceq^\flat, V^\flat\rangle$ from $\mathcal{M}$ and $\mathcal{O}$ in the same way as we defined $\mathcal{M}^\sharp$ from $\mathcal{M}$ and $\mathcal{N}$ for part 1, except that $\rightarrowtriangle^\flat = \rightarrowtriangle^{\mathcal{M}} \cup \rightarrowtriangle^{\mathcal{O}} \cup \{w, u\}$ (see Fig. 5). Observe that $\mathcal{M}^\flat, w$ is a linear pointed pre-model.

It is easy to verify that for all formulas ξ and $x \in M \setminus \{w\}$,

$$\mathcal{M}^\flat, x \vDash_d \xi \text{ iff } \mathcal{M}, x \vDash_d \xi \text{ ; and } \mathcal{M}^\flat, u \vDash_d \xi \text{ iff } \mathcal{O}, u \vDash_d \xi. \tag{5.22}$$

As in the proof of part 1, (5.19) holds for $\mathcal{M}$. It follows by the construction of $\mathcal{M}^\flat$ and (5.22) that (5.19) also holds for $\mathcal{M}^\flat$ and $\rightarrowtriangle^\flat$ in place of $\mathcal{M}$ and $\rightarrowtriangle^{\mathcal{M}}$, so $\mathcal{M}^\flat, w \vDash_d K\varphi_1 \wedge \cdots \wedge K\varphi_n$ by the truth definition. Also as in the proof of part 1, $\mathrm{Min}_{\preceq^{\mathcal{M}}_w}\left(\overline{[\![\psi_k]\!]}^{\mathcal{M}}\right) \cap \rightarrowtriangle^{\mathcal{M}}(w)$ is nonempty for all $k \leq m$. It follows by the construction of $\mathcal{M}^\flat$ and (5.22) that $\mathrm{Min}_{\preceq^\flat_w}\left(\overline{[\![\psi_k]\!]}^{\mathcal{M}^\flat}\right) \cap \rightarrowtriangle^\flat(w)$ is nonempty for all $k \leq m+1$, so $\mathcal{M}^\flat, w \nvDash K\psi_1 \vee \cdots \vee K\psi_{m+1}$ by the truth definition. Finally, since

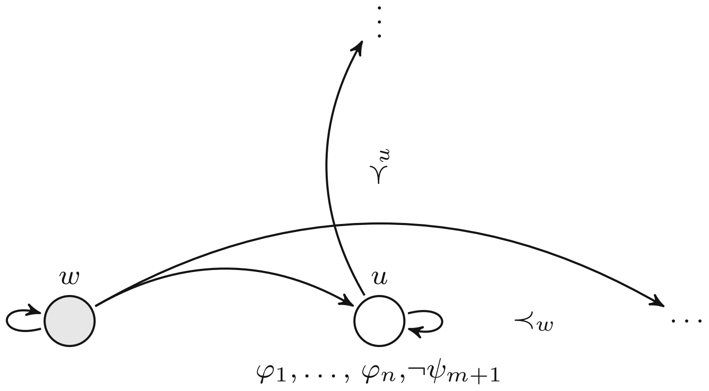

Fig. 5 Part of the extended pre-model $\mathcal{M}^\flat$ for Lemma 5.19.2

φ_0 is propositional, $\mathcal{M}, w \vDash \varphi_0$ implies $\mathcal{M}^\flat, w \vDash \varphi_0$ by definition of $V^\flat$. It follows from the preceding facts that $\mathcal{M}^\flat, u \nvDash_d \chi_{n,m+1}$. □

Remark 5.20 (*Properties of* $\rightarrowtriangle$) Lemma 5.19 also holds for the class of RA models/pre-models in which $\rightarrowtriangle$ is an equivalence relation, so that Theorem 5.2.2–3 will as well. For part 1, if $\mathcal{M}$ and $\mathcal{N}$ are in this class, so is $\mathcal{M}^\sharp$, since the union of two disjoint equivalence relations is an equivalence relation. For part 2, suppose $\mathcal{M}$ and $\mathcal{O}$ are in the class. Since we have added an arrow from w to u, $\mathcal{M}^\flat$ may not be in the class. In this case, let $\rightarrowtriangle^+$ be the minimal extension of $\rightarrowtriangle^\flat$ that is an equivalence relation. One can check that by construction of $\mathcal{M}^\flat$, for all $w \in W^\flat$,

$$(\rightarrowtriangle^+(w)\backslash \rightarrowtriangle^\flat(w)) \cap W_w = \emptyset.$$

It follows that $\mathcal{M}^\flat$ and $\mathcal{M}^+ = \langle W^\flat, \rightarrowtriangle^+, \preceq^\flat, V^\flat\rangle$ satisfy the same formulas according to D-semantics.

Using Lemma 5.19, we can now carry out our inductive construction.

Lemma 5.21 (**Pre-model Construction**) *If neither (a) nor (c) of Theorem 5.2 holds for $\chi_{n,m}$, then there is a linear pointed pre-model $\mathcal{M}, w$ such that $\mathcal{M}, w \nvDash_d \chi_{n,m}$.*

Proof The proof is by induction on m with a subsidiary induction on n.

Base Case for m Assume that neither (a) nor (c) holds for $\chi_{n,0}$.[43] Let $\mathcal{M} = \langle W, \rightarrowtriangle, \preceq, V\rangle$ be such that $W = \{w\}$, $\rightarrowtriangle = \{\langle w, w\rangle\}$, $\preceq_w = \emptyset$, and V is any valuation such that $\mathcal{M}, w \vDash \varphi_0$, which exists since (a) does not hold for $\chi_{n,0}$. Then $\mathcal{M}, w$ is a linear pointed pre-model such that $\mathcal{M}, w \nvDash_d \chi_{n,0}$.

Inductive Step for m Assume for induction on m that for any $\beta_1, \ldots, \beta_m$ and any n, if neither (a) nor (c) holds for $\chi := \varphi_0 \wedge K\varphi_1 \wedge \cdots \wedge K\varphi_n \rightarrow K\beta_1 \vee \cdots \vee K\beta_m$, then there is a linear pointed pre-model $\mathcal{M}, w$ with $\mathcal{M}, w \nvDash_d \chi$. Assume that for some $\psi_1, \ldots, \psi_{m+1}$, neither (a) nor (c) holds for $\chi_{n,m+1}$. We prove by induction on n that there a linear $\mathcal{M}', w$ with $\mathcal{M}', w \nvDash_d \chi_{n,m+1}$.

Base Case for n Assume neither (a) nor (c) holds for $\chi_{0,m+1}$. Since (c) does not hold, for all $j \leq m+1$, $\nvDash_d \top \leftrightarrow \psi_j$ and hence $\nvDash_d \top \rightarrow \psi_j$. Starting with $\mathcal{M}, w$ defined as in the base case for m such that $\mathcal{M}, w \nvDash \chi_{0,0}$, apply Lemma 5.19.2 $m+1$ times to obtain an $\mathcal{M}', w$ with $\mathcal{M}', w \nvDash \chi_{0,m+1}$.

Inductive Step for n Assume for induction on n that for any $\alpha_0, \ldots, \alpha_n$, if neither (a) nor (c) holds for $\chi := \alpha_0 \wedge K\alpha_1 \wedge \cdots \wedge K\alpha_n \rightarrow K\psi_1 \vee \cdots \vee K\psi_{m+1}$, then there is a linear pointed pre-model $\mathcal{M}, w$ with $\mathcal{M}, w \nvDash_d \chi$. Assume that for some $\varphi_0, \ldots, \varphi_{n+1}$, neither (a) nor (c) holds for $\chi_{n+1,m+1}$.

[43] Recall that $\chi_{n,0}$ is of the form $\varphi_0 \wedge K\varphi_1 \wedge \cdots \wedge K\varphi_n \rightarrow \bot$.

Case 1 $\vDash_d \varphi_1 \wedge \cdots \wedge \varphi_{n+1} \to \psi_1 \wedge \cdots \wedge \psi_{m+1}$. Then since (c) does not hold for $\chi_{n+1,m+1}$, $\nvDash_d \psi_1 \wedge \cdots \wedge \psi_{m+1} \to \varphi_1 \wedge \cdots \wedge \varphi_{n+1}$, in which case $\nvDash_d \psi_1 \wedge \cdots \wedge \psi_{m+1} \to \varphi_i$ for some $i \leq n+1$. Without loss of generality, assume

$$\nvDash_d \psi_1 \wedge \cdots \wedge \psi_{m+1} \to \varphi_{n+1}. \tag{5.23}$$

Since neither (a) nor (c) holds for $\chi_{n+1,m+1}$, neither holds for $\chi_{n,m+1}$. Hence by the inductive hypothesis for n there is a linear pointed pre-model $\mathcal{M}, w$ such that $\mathcal{M}, w \nvDash_d \chi_{n,m+1}$, which with (5.23) and Lemma 5.19.1 implies that there is a linear pointed pre-model $\mathcal{M}^\sharp, w$ such that $\mathcal{M}^\sharp, w \nvDash_d \chi_{n+1,m+1}$.

Case 2 $\nvDash_d \varphi_1 \wedge \cdots \wedge \varphi_{n+1} \to \psi_1 \wedge \cdots \wedge \psi_{m+1}$. Then for some $j \leq m+1$, $\nvDash \varphi_1 \wedge \cdots \wedge \varphi_{n+1} \to \psi_j$. Without loss of generality, assume

$$\nvDash_d \varphi_1 \wedge \cdots \wedge \varphi_{n+1} \to \psi_{m+1}. \tag{5.24}$$

Since neither (a) nor (c) holds for $\chi_{n+1,m+1}$, neither holds for $\chi_{n+1,m}$. Hence by the inductive hypothesis for m there is a linear pointed pre-model $\mathcal{M}, w$ such that $\mathcal{M}, w \nvDash_d \chi_{n+1,m}$, which with (5.24) and Lemma 5.19.2 implies that there is a linear pointed pre-model $\mathcal{M}^\flat, w$ such that $\mathcal{M}^\flat, w \nvDash_d \chi_{n+1,m+1}$. □

Finally, if $\chi_{n,m}$ is T-unpacked (Def. 5.15), then we can convert the falsifying pre-model obtained from Lemma 5.21 into a falsifying RA model.

Lemma 5.22 (**Pre-model to Model Conversion**) *Given a linear pointed pre-model $\mathcal{M}, w$ and a T-unpacked $\chi_{n,m}$ such that $\mathcal{M}, w \nvDash_d \chi_{n,m}$, there is a linear pointed RA model $\mathcal{M}^c, w$ such that $\mathcal{M}^c, w \nvDash_d \chi_{n,m}$.*

Proof Where $\mathcal{M} = \langle W, \multimap, \preceq, V\rangle$, define $\mathcal{M}^c = \langle W, \multimap, \preceq^c, V\rangle$ such that for all $v \in W \setminus \{w\}$, $\preceq^c_v = \preceq_v$, and $\preceq^c_w = \preceq_w \cup \{\langle w, v\rangle \mid v \in \{w\} \cup W_w\}$, where W_w is the field of $\preceq_w$. Since w is strictly minimal in $\preceq^c_w$, $\mathcal{M}^c$ is a linear RA model. (Note, however, that w is still not in the field of $\preceq^c_v$ for any $v \in W \setminus \{w\}$.) By construction of $\mathcal{M}^c$, together $\mathcal{M}, w \nvDash_d K\psi_1 \vee \cdots \vee K\psi_m$ and $w \multimap w$ imply

$$\mathcal{M}^c, w \nvDash_d K\psi_1 \vee \cdots \vee K\psi_m. \tag{5.25}$$

We prove by induction that for all $k \leq n$,

$$\mathcal{M}^c, w \vDash_d \varphi_0 \wedge K\varphi_1 \wedge \cdots \wedge K\varphi_k. \tag{5.26}$$

The base case of $k = 0$ is immediate since φ_0 is propositional, $\mathcal{M}, w \vDash \varphi_0$, and $\mathcal{M}$ and $\mathcal{M}^c$ have the same valuations. Assuming (5.26) holds for $k < n$, we must show $\mathcal{M}^c, w \vDash_d K\varphi_{k+1}$. Since $\chi_{n,m}$ is T-unpacked, together Definition 5.15, (5.25), and (5.26) imply $\mathcal{M}^c, w \vDash_d \varphi_{k+1}$. Since $\mathcal{M}, w \vDash_d K\varphi_{k+1}$, we have $\mathrm{Min}_{\preceq_w}(\overline{[\![\varphi_{k+1}]\!]}^{\mathcal{M}}) \cap \multimap(w) = \emptyset$ by the truth definition (Def. 3.6). It follows, given the construction of $\mathcal{M}^c$ and the fact that $\mathcal{M}^c, w \vDash_d \varphi_{k+1}$, that $\mathrm{Min}_{\preceq^c_w}(\overline{[\![\varphi_{k+1}]\!]}^{\mathcal{M}^c}) \cap \multimap(w) = \emptyset$, which gives $\mathcal{M}^c, w \vDash_d K\varphi_{k+1}$, as desired. □

The proof of the 'only if' direction of Theorem 5.2.2 is complete. By Lemmas 5.21 and 5.22, if a T-unpacked $\chi_{n,m}$ does not satisfy (a) or (c) of Theorem 5.2, then it is falsified by a linear—and hence total—RA model according to D-semantics. Indeed, as the next proposition and Corollary 5.24 together show, it is falsified by an RA model with the *universal field* property (Def. 3.3.4).

Proposition 5.23 (**Universalization**) *Where $\mathcal{M} = \langle W, \rightarrowtriangle, \preceq, V\rangle$ is a finite RA model, there is a finite RA model $\mathcal{M}^u = \langle W^u, \rightarrowtriangle^u, \preceq^u, V^u\rangle$ with the universal field property, such that $W \subseteq W^u$ and for all $w \in W$ and all φ,*

$$\mathcal{M}, w \vDash_d \varphi \text{ iff } \mathcal{M}^u, w \vDash_d \varphi.$$

If $\mathcal{M}$ is total, $\mathcal{M}^u$ is also total. If $\mathcal{M}$ is linear, $\mathcal{M}^u$ is also linear.

Proof Given $\mathcal{M} = \langle W, \rightarrowtriangle, \preceq, V\rangle$, suppose that for some $w, v \in W$, $v \notin W_w$, so $v \neq w$. Define $\mathcal{M}' = \langle W', \rightarrowtriangle', \preceq', V'\rangle$ such that $W' = W$; $\rightarrowtriangle' = \rightarrowtriangle \setminus\{\langle w, v\rangle\}$; $\preceq'_w = \preceq_w \cup \{\langle x, v\rangle \mid x \in W_w \cup \{v\}\}$; $\preceq'_y = \preceq_y$ for $y \in W \setminus \{w\}$; and $V' = V$. In other words, v becomes the least relevant world at w and eliminated at w in $\mathcal{M}'$. Given $v \notin W_w$, one can show by induction on φ that for all $x \in W$, $\mathcal{M}, x \vDash_d \varphi$ iff $\mathcal{M}', x \vDash_d \varphi$. Applying the transformation $\mathcal{M} \mapsto \mathcal{M}'$ successively no more than $|W|^2$ times with other pairs of worlds like w and v yields a model $\mathcal{M}^u$ with the universal field property. If $\mathcal{M}$ is total/linear, so is $\mathcal{M}^u$.

If we require that $\rightarrowtriangle$ be an equivalence relation, then the transformation above will not work in general, since we may lose transitivity or symmetry by setting $w \not\rightarrowtriangle' v$. To solve this problem, we first make an isomorphic copy of $\mathcal{M}$, labeled $\mathcal{M}^\star = \langle W^\star, \rightarrowtriangle^\star, \preceq^\star, V^\star\rangle$. For every $w \in W$, let $w^\star$ be its isomorphic copy in $W^\star$. Define $\mathcal{N} = \langle W^{\mathcal{N}}, \rightarrowtriangle^{\mathcal{N}}, \preceq^{\mathcal{N}}, V^{\mathcal{N}}\rangle$ as follows: $W^{\mathcal{N}} = W \cup W^\star$; $\rightarrowtriangle^{\mathcal{N}} = \rightarrowtriangle \cup \rightarrowtriangle^\star$; $V^{\mathcal{N}}(p) = V(p) \cup V^\star(p)$; for all $w \in W$, $\preceq^{\mathcal{N}}_w = \preceq_w \cup \{\langle v, u\rangle \mid v \in W^{\mathcal{N}}$ and $u \in W^\star\}$; for all $w^\star \in W^\star$, $\preceq^{\mathcal{N}}_{w^\star} = \preceq_{w^\star} \cup \{\langle v, u\rangle \mid v \in W^{\mathcal{N}}$ and $u \in W\}$. In other words, $\mathcal{N}$ is the result of first taking the disjoint union of $\mathcal{M}$ and $\mathcal{M}^\star$ (so there are no $v \in W$ and $u \in W^\star$ such that $v \rightarrowtriangle^{\mathcal{N}} u$ or $u \rightarrowtriangle^{\mathcal{N}} v$) and then making all worlds in $W^\star$ the least relevant worlds from the perspective of all worlds in W, and vice versa.[44] Given this construction, it is easy to prove by induction that for all $w \in W$ and formulas φ, $\mathcal{M}, w \vDash_d \varphi$ iff $\mathcal{N}, w \vDash_d \varphi$ iff $\mathcal{N}, w^\star \vDash_d \varphi$. Moreover, $\rightarrowtriangle^{\mathcal{N}}$ is an equivalence relation if $\rightarrowtriangle$ is.

Next we turn $\mathcal{N}$ into a model with universal fields, without changing $\rightarrowtriangle^{\mathcal{N}}$. Suppose that for $w, v \in W$, v is not in the field of $\preceq^{\mathcal{N}}_w$, which is the case iff $v^\star$ is not in the field of $\preceq^{\mathcal{N}}_{w^\star}$. (Remember that for all $w \in W$ and $u \in W^\star$, u is in the field of $\preceq^{\mathcal{N}}_w$ and vice versa). Let $\mathcal{N}' = \langle W', \rightarrowtriangle', \preceq', V'\rangle$ be such that: $W' = W^{\mathcal{N}}$; $\rightarrowtriangle' = \rightarrowtriangle^{\mathcal{N}}$;

[44] If we want to stay within the class of *linear* models, then we must change the definition of $\preceq^{\mathcal{N}}_w$ so that it extends the linear order $\preceq_w$ with an arbitrary linear order on $W^\star$ that makes all worlds in $W^\star$ less relevant than all worlds in W, and similarly for $\preceq^{\mathcal{N}}_{w^\star}$.

$V' = V^{\mathcal{N}}$; for all $u \in W' \setminus \{w, w^\star\}$, $\preceq'_u = \preceq^{\mathcal{N}}_u$; $\preceq'_w = \preceq^{\mathcal{N}}_w \cup \{\langle x, v \rangle \mid x \in W^{\mathcal{N}}_w \cup \{v\}\}$; and $\preceq'_{w^\star} = \preceq^{\mathcal{N}}_{w^\star} \cup \{\langle x, v^\star \rangle \mid x \in W^{\mathcal{N}}_{w^\star} \cup \{v^\star\}\}$. It follows that for all $x \in W^{\mathcal{N}}_w$, $x \preceq'_w v^\star \prec'_w v$; and for all $x \in W^{\mathcal{N}}_{w^\star}$, $x \preceq'_{w^\star} v \prec'_{w^\star} v^\star$. Since $w \not\rhd' v^\star$ and $w^\star \not\rhd' v$, one can prove by induction that for all φ and $u \in W$, $\mathcal{N}, u \vDash_d \varphi$ iff $\mathcal{N}', u \vDash_d \varphi$ iff $\mathcal{N}', u^\star \vDash_d \varphi$. The key is that although we put v in the field of $\preceq'_w$, this cannot make any $K\psi$ formula that is true at $\mathcal{N}, w$ false at $\mathcal{N}', w$, for if $\mathcal{N}', v \nvDash_d \psi$, then by the inductive hypothesis $\mathcal{N}', v^\star \nvDash_d \psi$, and $v^\star$ is more relevant than v and eliminated at w; similarly, although we put $v^\star$ in the field of $\preceq'_{w^\star}$, this cannot make any $K\psi$ formula that is true at $\mathcal{N}, w^\star$ false at $\mathcal{N}', w^\star$. Applying the transformation $\mathcal{N} \mapsto \mathcal{N}'$ successively no more than $|W^{\mathcal{N}}|^2$ times with other worlds like w and v yields a universalized $\mathcal{M}^u$. □

5.2.3 Finite Models and Complexity

From the proofs of Section 5.2.2, we obtain results on finite models and the complexity of satisfiability for D-semantics over total (linear, universal) RA models.

Corollary 5.24 (**Effective Finite Model Property**) *For any formula φ of the epistemic language, if φ is satisfiable in a total RA model according to D-semantics, then φ is satisfiable in a total RA model $\mathcal{M}$ with $|\mathcal{M}| \leq |\varphi|^{d(\varphi)}$.*

Proof By strong induction on $d(\varphi)$. Since φ is satisfiable iff $\neg\varphi$ is falsifiable, consider the latter. By Proposition 5.16, $\neg\varphi$ is equivalent to a conjunction of T-unpacked formulas of the form $\chi_{n,m}$, which is falsifiable iff one of its conjuncts $\chi_{n,m}$ is falsifiable. By Lemmas 5.19–5.22, if $\chi_{n,m}$ is falsifiable, then it is falsifiable in a model $\mathcal{M}$ that combines at most k other models (and one root world), where k is the number of top-level K operators in $\chi_{n,m}$, which is bounded by $|\varphi|$. Each of the these models is selected as a model of a formula of lesser modal depth than $\chi_{n,m}$, so by the inductive hypothesis we can assume that each is of size at most $|\varphi|^{d(\varphi)-1}$. Hence $|\mathcal{M}| \leq |\varphi| \times |\varphi|^{d(\varphi)-1} = |\varphi|^{d(\varphi)}$. □

Corollary 5.25 (**Complexity of Satisfiability**)

1. *The problem of deciding whether an epistemic formula is satisfiable in the class of total RA models according to D-semantics is in PSPACE;*
2. *For any k, the problem of deciding whether an epistemic formula φ with $d(\varphi) \leq k$ is satisfiable in the class of total RA models according to D-semantics is NP-complete.*

Proof (Sketch) For part 1, given PSPACE = NPSPACE (see Papadimitriou [59, Section 7.3]), it suffices to give a non-deterministic algorithm using polynomial space. By the previous results (including Prop. 5.16), if φ is satisfiable, then it is satisfiable in a model that can be inductively constructed as in the proofs of Lemmas 5.19, 5.21, and 5.22. We want an algorithm to non-deterministically guess such a model. However, since the size of the model may be exponential in $|\varphi|$, we cannot necessarily store the entire model in memory using only polynomial space. Instead, we non-deterministically guess the submodels that are combined in the inductive construction, taking advantage of the following fact from the proof of Lemma 5.19. Once we have computed the truth values at $\mathcal{N}, v$ (or $\mathcal{O}, u$) of all subformulas of φ (up to some modal depth, depending on the stage of the construction), we can label v with the true subformulas and then erase the rest of $\mathcal{N}$ from memory (and similarly for $\mathcal{O}, u$). The other worlds in $\mathcal{N}$ will not be in the field of $\preceq_x$ for any world x at which we need to compute truth values at any later stage of the construction, so it is not necessary to access those worlds in order to compute later truth values. Given this space-saving method, we only need to use polynomial space at any given stage of the algorithm. I leave the details of the algorithm to the reader.[45]

For part 2, NP-hardness is immediate, since for $k = 0$ we have all formulas of propositional logic. For membership in NP, if φ is satisfiable and $d(\varphi) \leq k$, then by Corollary 5.24, φ satisfiable in a model $\mathcal{M}$ with $|\mathcal{M}| \leq |\varphi|^k$. We can non-deterministically guess such a model, and it is easy to check that evaluating φ in $\mathcal{M}$ is in polynomial time given that $\mathcal{M}$ is polynomial-sized. □

As explained in Remark 7.2, Corollary 5.25.1 accords with results of Vardi [72]. Corollary 5.25.2 accords with results of Halpern [28] on the effect of bounding modal depth on the complexity of satisfiability for modal logics.

5.3 Completeness for All RA Models

Next we prove the 'only if' direction of Theorem 5.2.3. In the process we prove the separation property for D-semantics over all RA models noted in Proposition 5.6. Interestingly, dropping totality makes things simpler.

Claim If neither (a) nor (d) holds for a T-unpacked $\chi_{n,m}$, then there is a pointed RA model $\mathcal{M}, w$ such that $\mathcal{M}, w \nvDash_d \chi_{n,m}$.

Proof If $m \leq 1$, (d) is the same as (c), covered in Section 5.2.2. So suppose $m > 1$. By Lemma 5.22 and the $m = 1$ case of the inductive proof of Lemma 5.21, if neither (a) nor (d) holds for $\chi_{n,m}$, then for all $1 \leq j \leq m$, there is a linear RA model $\mathcal{M}_j = \langle W_j, \rightarrowtriangle_j, \preceq^j, V_j \rangle$ with point $w_j \in W_j$ such that

$$\mathcal{M}_j, w_j \nvDash_d \varphi_0 \wedge K\varphi_1 \wedge \cdots \wedge K\varphi_n \rightarrow K\psi_j. \tag{5.27}$$

Recall that $\mathcal{M}_j$ is constructed in such a way that for all $v \in W_j^- = W_j \setminus \{w_j\}$, w_j is not in the field of $\preceq^j_v$. Without loss of generality, assume that for all $j, k \leq m$, $W_j \cap W_k = \emptyset$. Construct $\mathcal{M} = \langle W, \rightarrowtriangle, \preceq, V \rangle$ as follows, by first taking the disjoint union of all of the $\mathcal{M}_j$, then "merging" all of the w_j into a single new world w (with

[45] Cf. Theorem 4.2 of Friedman and Halpern [24] for a proof that the complexity of satisfiability for formulas of conditional logic in similar preorder structures is in PSPACE.

the same valuation as some w_k), so that the linear models $\mathcal{M}_j$ are linked to w like spokes to the hub of a wheel (recall Fig. 3):

$$W = \{w\} \cup \bigcup_{j\leq m} W_j^- ;\text{ for all } j \leq m \text{ and } v \in W_j^-, \preceq_v = \preceq_v^j;$$

$$\preceq_w = \{\langle w, v\rangle \mid v = w \text{ or } \exists\, j \leq m\colon w_j \preceq_{w_j}^j v\} \cup \bigcup_{j\leq m} (\preceq_{w_j}^j \cap (W_j^- \times W_j^-));$$

$$\rightarrowtriangle\, = \{\langle w, v\rangle \mid v = w \text{ or } \exists\, j \leq m\colon w_j \rightarrowtriangle_j v\} \cup \bigcup_{j\leq m} (\rightarrowtriangle_j \cap (W_j^- \times W_j^-));$$

$$V(p) = \begin{cases} \bigcup_{j\leq m} (V_j(p) \cap W_j^-) \cup \{w\} & \text{if } w_1 \in V_1(p); \\ \bigcup_{j\leq m} (V_j(p) \cap W_j^-) & \text{if } w_1 \notin V_1(p). \end{cases}$$

It is easy to verify that for all formulas ξ, $j \leq m$, and $v \in W_j^-$,

$$\mathcal{M}, v \vDash_d \xi \text{ iff } \mathcal{M}_j, v \vDash_d \xi. \tag{5.28}$$

It follows from the construction of $\mathcal{M}$ and (5.28) that for all $j \leq m$,

$$\mathrm{Min}_{\preceq_{w_j}^j}\left(\overline{[\![\psi_j]\!]}^{\mathcal{M}_j}\right) \cap \rightarrowtriangle_j(w) \subseteq \mathrm{Min}_{\preceq_w}\left(\overline{[\![\psi_j]\!]}^{\mathcal{M}}\right) \cap \rightarrowtriangle(w). \tag{5.29}$$

For all $j \leq m$, given $\mathcal{M}_j, w_j \nvDash_d K\psi_j$ by assumption, the left side of (5.29) is nonempty, so the right side is nonempty. Hence by the truth definition,

$$\mathcal{M}, w \nvDash_d K\psi_1 \vee \cdots \vee K\psi_m. \tag{5.30}$$

By our initial assumption, for all $j \leq m$,

$$\bigcup_{i\leq n} \mathrm{Min}_{\preceq_{w_j}^j}\left(\overline{[\![\varphi_i]\!]}^{\mathcal{M}_j}\right) \cap \rightarrowtriangle^j(w) = \emptyset. \tag{5.31}$$

We prove by induction that for $1 \leq i \leq n$,

$$\mathrm{Min}_{\preceq_w}\left(\overline{[\![\varphi_i]\!]}^{\mathcal{M}}\right) \cap \rightarrowtriangle(w) = \emptyset. \tag{5.32}$$

Base Case Given $\mathcal{M}_1, w_1 \vDash \varphi_0$ and the fact that w has the same valuation under V as w_1 under V_1, we have $\mathcal{M}, w \vDash \varphi_0$. Together with (5.30), this implies $\mathcal{M}, w \nvDash_d \chi_{0,m}$. Since $\chi_{1,m}$ is T-unpacked, it follows by Definition 5.15 that $\mathcal{M}, w \vDash_d \varphi_1$, in which case $w \notin \mathrm{Min}_{\preceq_w}\left(\overline{[\![\varphi_1]\!]}^{\mathcal{M}}\right)$. By construction of $\mathcal{M}$, together (5.31), (5.28), and $w \notin \mathrm{Min}_{\preceq_w}\left(\overline{[\![\varphi_1]\!]}^{\mathcal{M}}\right)$ imply (5.32) for $i = 1$.

Inductive Step Assume (5.32) for all $k \leq i$ $(i < n)$, so $\mathcal{M}, w \vDash_d K\varphi_1 \wedge \cdots \wedge K\varphi_i$, which with (5.30) gives $\mathcal{M}, w \nvDash_d \chi_{i,m}$. Then since $\chi_{i+1,m}$ is T-unpacked, $\mathcal{M}, w \vDash_d \varphi_{i+1}$, so by reasoning as in the base case, (5.32) holds for $i + 1$.

Since (5.32) holds for $1 \leq i \leq n$, by the truth definition we have $\mathcal{M}, w \vDash_d K\varphi_1 \wedge \cdots \wedge K\varphi_n$, which with $\mathcal{M}, w \vDash \varphi_0$ and (5.30) implies $\mathcal{M}, w \nvDash_d \chi_{n,m}$. □

A remark analogous to Remark 5.20 applies to the above construction: if each $\twoheadrightarrow_j$ is an equivalence relation and we extend $\twoheadrightarrow$ to the minimal equivalence relation $\twoheadrightarrow^+ \supseteq \twoheadrightarrow$, then the resulting model will still falsify $\chi_{n,m}$. Hence Theorem 5.2.3 holds for the class of RA models with equivalence relations (and with the universal field property by Prop. 5.23). Finally, arguments similar to those of Corollaries 5.24–5.25 show the finite model property and PSPACE satisfiability without the assumption of totality (see Remark 7.2).

5.4 Completeness for CB Models

Finally, for the 'only if' direction of Theorem 5.2.4, there are two ways to try to falsify some $\chi_{n,m}$. For H/N-semantics, we can first construct an RA countermodel for $\chi_{n,m}$ under D-semantics, as in Section 5.2, and then transform it into a CB countermodel for $\chi_{n,m}$ under H/N-semantics, as shown in Section 6 below. Alternatively, we can first construct a CB countermodel under S/H-semantics and then transform it into a CB countermodel under H/N-semantics as in Section 6. Here we will take the latter route. By Proposition 6.2 below, for the 'only if' direction of Theorem 5.2.4 it suffices to prove the following.

Claim If neither (a) nor (d) holds for a flat, T-unpacked $\chi_{n,m}$, then there is a pointed CB model $\mathcal{M}, w$ such that $\mathcal{M}, w \nvDash_{h,s} \chi_{n,m}$.

We begin with some notation used in the proof and in later sections.

Notation 5.26 (**Relational Image**) Given a CB model $\mathcal{M} = \langle W, D, \leqslant, V\rangle$, the image of $\{w\}$ under the relation D is $D(w) = \{v \in W \mid wDv\}$.

Hence $D(w)$ is the set of doxastically accessible worlds for the agent in w.

Let us now prove the claim.

Proof For any positive integer z, let $P_z = \{1, \ldots, z\}$. For all $k \in P_m$, let $S_k = \{i \in P_n \mid \vDash \psi_k \to \varphi_i\}$, and $T = \{t \in P_m \mid S_t = P_n\}$. Since (d) does not hold for $\chi_{n,m}$, it follows that

$$\nvDash \bigwedge_{i \in S_k} \varphi_i \to \psi_k. \tag{5.33}$$

Construct $\mathcal{M} = \langle W, D, \leqslant, V\rangle$ as follows (see Fig. 6):

$W = \{w\} \cup \{x_t \mid t \in T\} \cup \{v_k, u^k_j \mid k \in P_m \setminus T \text{ and } j \in P_n \setminus S_k\}$;

D is the union of $\{\langle w, w\rangle\}$, $\{\langle w, x_t\rangle, \langle x_t, x_t\rangle \mid t \in T\}$, and

$$\left\{\langle v_k, u^k_j\rangle, \langle u^k_j, u^k_j\rangle \mid k \in P_m \setminus T \text{ and } j \in P_n \setminus S_k\right\};$$

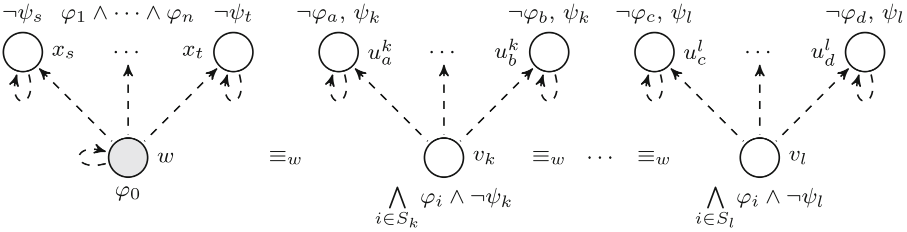

Fig. 6 Countermodel for $\chi_{n,m}$ in H/S-semantics

$\leqslant_w = \{\langle w, w\rangle\} \cup \{\langle w, v_k\rangle, \langle v_k, w\rangle, \langle v_k, v_k\rangle \mid k \in P_m\}$;[46]
For $y \in W \setminus \{w\}$, $\leqslant_y$ is any relation as in Definition 3.1.3;
V is any valuation function on W such that $\mathcal{M}, w \vDash \varphi_0$ and

- for all $t \in T$, $\mathcal{M}, x_t \vDash \bigwedge_{i\in P_n} \varphi_i \wedge \neg\psi_t$;
- for all $k \in P_m \setminus T$, $\mathcal{M}, v_k \vDash \bigwedge_{i\in S_k} \varphi_i \wedge \neg\psi_k$;
- for all $k \in P_m \setminus T$ and $j \in P_n \setminus S_k$, $\mathcal{M}, u_j^k \vDash \neg\varphi_j \wedge \psi_k$.

Such a valuation V exists by the assumption that (a) does not hold for $\chi_{n,m}$, together with (5.33) and the definitions of T and S_k.

Since $\chi_{n,m}$ is flat and T-unpacked, $\mathcal{M}, w \vDash \varphi_0$ implies $\mathcal{M}, w \vDash \varphi_1 \wedge \cdots \wedge \varphi_n$. Then since $D(w) = \{w\} \cup \{x_t \mid t \in T\}$ and $\mathcal{M}, x_t \vDash \varphi_1 \wedge \cdots \wedge \varphi_n$ for all $t \in T$,

$$\mathcal{M}, w \vDash \bigwedge_{i\in P_n} (B\varphi_i \wedge \varphi_i). \tag{5.34}$$

For all $k \in P_m \setminus T$, we have

$$\mathcal{M}, v_k \nvDash \bigvee_{j\in P_n \setminus S_k} B\varphi_j \tag{5.35}$$

given $v_k D u_j^k$ and $\mathcal{M}, u_j^k \nvDash \varphi_j$, and

$$\mathcal{M}, v_k \vDash \bigwedge_{i\in S_k} \varphi_i \tag{5.36}$$

by definition of V. It follows from (5.35) and (5.36) that for all $k \in P_m \setminus T$,

$$\mathcal{M}, v_k \vDash \bigwedge_{i\in P_n} (B\varphi_i \to \varphi_i). \tag{5.37}$$

By construction of $\mathcal{M}$, (5.34) and (5.37) together imply that for all $y \in W_w$,

$$\mathcal{M}, y \vDash \bigwedge_{i\in P_n} (B\varphi_i \to \varphi_i). \tag{5.38}$$

[46] The x_t and u_j^k worlds are not in the field of $\leqslant_w$. For a universal field (and total relation), the proof works with minor additions if we take the union of $\leqslant_w$ as defined above with

$$\{\langle w, x_t\rangle, \langle w, u_j^k\rangle, \langle v_k, x_t\rangle, \langle v_k, u_j^k\rangle, \langle x_t, x_t\rangle, \langle x_t, u_j^k\rangle, \langle u_j^k, u_j^k\rangle \mid t \in T, k \in P_m \setminus T, j \in P_n \setminus S_k\}.$$

Together (5.34) and (5.38) imply $\mathcal{M}, w \vDash_{h,s} K\varphi_i$ for all $i \in P_n$ by the truth definitions (Def. 4.3). Now let us check that $\mathcal{M}, w \nvDash_{h,s} K\psi_i$ for all $i \in P_m$. On the one hand, for all $t \in T$, given wDx_t and $\mathcal{M}, x_t \nvDash \psi_t$, we have $\mathcal{M}, w \nvDash B\psi_t$ and hence $\mathcal{M}, w \nvDash_{h,s} K\psi_t$. On the other hand, for all $k \in P_m \setminus T$, given $D(v_k) = \{u_k^j \mid j \in P_n \setminus S_k\}$ and $\mathcal{M}, u_j^k \vDash \psi_k$, we have $\mathcal{M}, v_k \vDash B\psi_k$; but then since $\mathcal{M}, v_k \nvDash \psi_k$ and $v_k \in \text{Min}_{\leqslant_w}(W)$, it follows that $\mathcal{M}, w \nvDash_{h,s} K\psi_k$. Together with $\mathcal{M}, w \vDash \varphi_0$, the previous facts imply $\mathcal{M}, w \nvDash_{h,s} \chi_{n,m}$. □

We leave the extension of the 'only if' direction of Theorem 5.2.4 to the full epistemic language for other work (see Problem 8.12). Facts 8.8.4, 8.8.5, and 8.10.1 show that for the full language, this direction must be modified. Yet for our purposes here, the above proof already helps to reveal the sources of closure failure in H/S-semantics and in N-semantics by Proposition 6.2 below.

5.5 The Sources of Closure Failure

The results of Sections 5.2–5.4 allow us to clearly identify the sources of closure failure in D/H/N/S-semantics. In D-semantics, the source of closure failure is the orderings—if we collapse the orderings, then D- is equivalent to L-semantics (see Observation 8.3) and closure failures disappear. By Proposition 6.1 below, the orderings are also a source of closure failure in H/N-semantics. However, the proof in Section 5.4 shows that there is another source of closure failure in H/N/S-semantics: the interpretation of ruling out in terms *belief*, as in the quote from Heller in Section 3. This is the sole source of closure failure in S-semantics, the odd member of the D/H/N/S-family that does not use the orderings beyond $\text{Min}_{\leqslant_w}(W)$ (recall Observation 4.5). Given this source of closure failure, even if we collapse the orderings, in which case H- is equivalent to S-semantics (see Prop. 6.3), closure failure persists. We will return to this point in Section 9.

6 Relating RA and CB Models

The discussion in Sections 5.4 and 5.5 appealed to claims about the relations between D/H/N/S-semantics. In this short section, we prove these claims. Readers eager to see how the results of Section 5 lead to complete deductive systems for the RA and subjunctivist theories should skip ahead to Section 7 and return here later.

One way to see how the RA and subjunctivist theories are related is by transforming models viewed from the perspective of one theory into models that are equivalent, with respect to what can be expressed in our language, when viewed from the perspective of another theory. This also shows that any closure principle that fails for the first theory also fails for the second.

We first see how to transform any RA model viewed from the perspective of D-semantics into a CB model that is equivalent, with respect to the flat fragment of the epistemic language, when viewed from the perspective of H-semantics. The transformation is intuitive: if, in the RA model, a possibility v is *eliminated* by the agent in w, then we construct the CB model such that if the agent were in situation v

instead of w, the agent would *notice*, i.e., would correctly believe that the true situation is v rather than w;[47] but if, in the RA model, v is *uneliminated* by the agent in w, then we construct the CB model such that if the agent were in situation v instead of w, the agent would *not* notice, i.e., would incorrectly believe that the true situation is w rather than v. (The CB model in Fig. 2 is obtained from the RA model in Fig. 1 in this way). Then the agent has eliminated the relevant alternatives to a flat φ at w in the RA model iff the agent *sensitively* believes φ at w in the CB model.

Proposition 6.1 (**D-to-H Transform**) *For any RA model* $\mathcal{M} = \langle W, \rightarrowtriangle, \preceq, V\rangle$ *with* $w \in W$*, there is a CB model* $\mathcal{N} = \langle W, D, \leqslant, V\rangle$ *such that for all flat epistemic formulas* φ*,*

$$\mathcal{M}, w \vDash_d \varphi \text{ iff } \mathcal{N}, w \vDash_h \varphi.$$

Proof Construct $\mathcal{N}$ from $\mathcal{M}$ as follows. Let W and V in $\mathcal{N}$ be the same as in $\mathcal{M}$; let $\leqslant$ in $\mathcal{N}$ be the same as $\preceq$ in $\mathcal{M}$; construct D in $\mathcal{N}$ from $\rightarrowtriangle$ in $\mathcal{M}$ as follows, where w is the fixed world in the lemma (recall Notation 5.26):

$$\forall v \in W\colon D(v) = \begin{cases} \{w\} \text{ if } w \rightarrowtriangle v; \\ \{v\} \ \text{ if } w \not\rightarrowtriangle v. \end{cases} \tag{6.1}$$

To prove the 'iff' by induction on φ, the base case is immediate and the boolean cases routine. Suppose φ is of the form $K\psi$. Since φ is flat, ψ is propositional. Given that V is the same in $\mathcal{N}$ as in $\mathcal{M}$, for all $v \in W$, $\mathcal{M}, v \vDash_d \psi$ iff $\mathcal{N}, v \vDash_h \psi$. Hence if $\mathcal{M}, w \nvDash_d \psi$, then $\mathcal{M}, w \nvDash_d K\psi$ and $\mathcal{N}, w \nvDash_h K\psi$ by Facts 3.7 and 4.4. Suppose $\mathcal{M}, w \vDash_d \psi$. Since $w \rightarrowtriangle w$, we have $D(w) = \{w\}$ by construction of $\mathcal{N}$, so $\mathcal{N}, w \vDash_h B\psi$ given $\mathcal{N}, w \vDash_h \psi$. It only remains to show that $\mathcal{M}, w \vDash_d K\psi$ iff the sensitivity condition (Def. 4.3) for $K\psi$ is satisfied at $\mathcal{N}, w$. This is easily seen to be a consequence of the following, given by the construction of $\mathcal{N}$:

$$\mathrm{Min}_{\preceq_w}\left(\overline{[\![\psi]\!]_d^{\mathcal{M}}}\right) = \mathrm{Min}_{\leqslant_w}\left(\overline{[\![\psi]\!]_h^{\mathcal{N}}}\right); \tag{6.2}$$

$$\forall u \in \mathrm{Min}_{\preceq_w}\left(\overline{[\![\psi]\!]_d^{\mathcal{M}}}\right)\colon w \rightarrowtriangle u \text{ iff } \mathcal{N}, u \vDash_h B\psi. \tag{6.3}$$

The left-to-right direction of the biconditional in (6.3) follows from the fact that if $w \rightarrowtriangle u$, then $D(u) = \{w\}$, and $\mathcal{N}, w \vDash_h \psi$. For the right-to-left direction, if $w \not\rightarrowtriangle u$, then $D(u) = \{u\}$, in which case $\mathcal{N}, u \nvDash_h B\psi$ given $\mathcal{N}, u \nvDash_h \psi$. □

The transformation above does not always preserve all non-flat epistemic formulas, and by Fact 8.8.4, no transformation does so. However, since the flat fragment of the language suffices to express all principles of closure with respect to propositional logic, Proposition 6.1 has the notable corollary that all such closure principles that fail in D-semantics also fail in H-semantics.

Next we transform CB models viewed from the perspective of H-semantics into CB models that are equivalent, with respect to the epistemic-doxastic language, when

[47] In fact, we only need something weaker, namely, that it would be *compatible* with what the agent believes that the true situation is v, i.e., vDv. In the $w \not\rightarrowtriangle v$ case of the definition of D in the proof of Proposition 6.1, we only need that $v \in D(v)$ for the proof to work.

viewed from the perspective of N-semantics. (Fact 8.8 in Section 8 shows that there is no such general transformation in the N-to-H direction.) To do so, we make the models *centered*, which (as noted in Observation 4.5) trivializes the adherence condition that separates N- from H-semantics.

Proposition 6.2 (**H-to-N Transform**) *For any CB model* $\mathcal{N} = \langle W, D, \leqslant, V\rangle$*, there is a CB model* $\mathcal{N}' = \left\langle W, D, \leqslant', V\right\rangle$ *such that for all* $w \in W$ *and all epistemic-doxastic formulas* φ*,*

$$\mathcal{N}, w \vDash_h \varphi \text{ iff } \mathcal{N}', w \vDash_n \varphi.$$

Proof Construct $\mathcal{N}'$ from $\mathcal{N}$ as follows. Let W, D, and V in $\mathcal{N}'$ be the same as in $\mathcal{N}$. For all $w \in W$, construct $\leqslant'_w$ from $\leqslant_w$ by making w strictly minimal in $\leqslant'_w$, but changing nothing else:

$$u \leqslant'_w v \text{ iff} \begin{cases} v \neq w \text{ and } u \leqslant_w v, \text{ or} \\ u = w. \end{cases} \tag{6.4}$$

To prove the proposition by induction on φ, the base case is immediate and the boolean and belief cases routine. Suppose φ is $K\psi$ and $[\![\psi]\!]_h^{\mathcal{N}} = [\![\psi]\!]_n^{\mathcal{N}'}$. If $\mathcal{N}, w \nvDash_h \psi$, then $\mathcal{N}, w \nvDash_h K\psi$ and $\mathcal{N}', w \nvDash_n K\psi$ by Fact 4.4. If $\mathcal{N}, w \vDash_h \psi$ and hence $\mathcal{N}', w \vDash_n \psi$, then by construction of $\leqslant'_w$ and the inductive hypothesis,

$$\mathrm{Min}_{\leqslant_w}\left(\overline{[\![\psi]\!]_h^{\mathcal{N}}}\right) = \mathrm{Min}_{\leqslant'_w}\left(\overline{[\![\psi]\!]_n^{\mathcal{N}'}}\right). \tag{6.5}$$

Since D is the same in $\mathcal{N}$ as in $\mathcal{N}'$, (6.5) implies that the belief and sensitivity conditions for $K\psi$ are satisfied at $\mathcal{N}, w$ iff they are satisfied at $\mathcal{N}', w$. If the belief condition is satisfied, then $\mathrm{Min}_{\leqslant'_w}([\![B\psi]\!]_n^{\mathcal{N}'}) = \{w\}$ by construction of $\leqslant'_w$, so the adherence condition (Def. 4.3) is automatically satisfied at $\mathcal{N}', w$. Hence the belief and sensitivity conditions for $K\psi$ are satisfied at $\mathcal{N}, w$ iff the belief, sensitivity, and adherence conditions are satisfied for $K\psi$ at $\mathcal{N}', w$.[48] □

Our last transformation takes us from models viewed from the perspective of S-semantics to equivalent models viewed from the perspective of H-semantics—and hence N-semantics by Proposition 6.2. (Fact 8.10 in Section 8 shows that there can be no such general transformation in the H-to-S direction.) The idea of the transformation is that safety is the ∃∀ condition (as in Section 3) obtained by restricting the scope of sensitivity to a fixed set of worlds, $\mathrm{Min}_{\leqslant_w}(W)$.

Proposition 6.3 (**S-to-H Transform**) *For any CB model* $\mathcal{N} = \langle W, D, \leqslant, V\rangle$*, there is a CB model* $\mathcal{N}' = \left\langle W, D, \leqslant', V\right\rangle$ *such that for all* $w \in W$ *and all epistemic-doxastic formulas* φ*,*

$$\mathcal{N}, w \vDash_s \varphi \text{ iff } \mathcal{N}', w \vDash_h \varphi.$$

[48] It is easy to see that even if we forbid centered models, Proposition 6.2 will still hold. For we can allow any number of worlds in $\mathrm{Min}_{\leqslant'_w}(W)$, provided they do not witness a violation of the adherence condition at w for any φ for which we want $\mathcal{N}, w \vDash_n K\varphi$.

Table 1 Axiom schemas and rules

PL. all tautologies	MP. $\dfrac{\varphi \to \psi \quad \varphi}{\psi}$
T. $K\varphi \to \varphi$ N. $K\top$	RE. $\dfrac{\varphi \leftrightarrow \psi}{K\varphi \leftrightarrow K\psi}$
M. $K(\varphi \wedge \psi) \to K\varphi \wedge K\psi$	RK. $\dfrac{\varphi_1 \wedge \cdots \wedge \varphi_n \to \psi}{K\varphi_1 \wedge \cdots \wedge K\varphi_n \to K\psi}\ (n \geq 0)$
X. $K(\varphi \wedge \psi) \to K\varphi \vee K\psi$	RAT. $\dfrac{\varphi_1 \wedge \cdots \wedge \varphi_n \leftrightarrow \psi_1 \wedge \cdots \wedge \psi_m}{K\varphi_1 \wedge \cdots \wedge K\varphi_n \to K\psi_1 \vee \cdots \vee K\psi_m}\ (n \geq 0, m \geq 1)$
C. $K\varphi \wedge K\psi \to K(\varphi \wedge \psi)$	RA. $\dfrac{\varphi_1 \wedge \cdots \wedge \varphi_n \leftrightarrow \psi}{K\varphi_1 \wedge \cdots \wedge K\varphi_n \to K\psi}\ (n \geq 0)$

Proof Construct $\mathcal{N}'$ from $\mathcal{N}$ as follows. Let W, D, and V in $\mathcal{N}'$ be the same as in $\mathcal{N}$. For all $w \in W$, construct $\leqslant'_w$ from $\leqslant_w$ by taking $\mathrm{Min}_{\leqslant_w}(W)$ to be the field of $\leqslant'_w$ and setting $u \leqslant'_w v$ for all u and v in the field. It is straightforward to check that $\mathcal{N}$ and $\mathcal{N}'$ are equivalent with respect to the safety condition and that in $\mathcal{N}'$ the safety and sensitivity conditions become equivalent.[49] □

Although I have introduced the propositions above for the purpose of relating the (in)valid closure principles of one theory to those of another, by transforming countermodels of one kind into countermodels of another, the interest of this style of analysis is not just in transferring principles for reasoning about knowledge between theories; the interest is also in highlighting the structural relations between different pictures of what knowledge is. In part II, we will continue our model-theoretic analysis to illuminate these pictures.

7 Deductive Systems

From Theorem 5.2 we obtain complete deductive systems for reasoning about knowledge according to the RA, tracking, and safety theories. Table 1 lists all of the needed schemas and rules, using the nomenclature of Chellas [12] (except for X, RAT, and RA, which are new). **E** is the weakest of the *classical* modal systems with PL, MP, and RE. $\mathbf{ES}_1 \ldots \mathbf{S}_n$ is the extension of **E** with every instance of schemas $\mathrm{S}_1 \ldots \mathrm{S}_n$. **EMCN** is familiar as the weakest normal modal system **K**, equivalently characterized in terms of PL, MP, the K schema, and the necessitation rule for K (even more simply, by PL, MP, and RK).

[49] It is easy to see that even if we require $W_w \setminus \mathrm{Min}_{\leqslant'_w}(W) \neq \emptyset$, Proposition 6.2 will still hold. For we can allow any number of worlds in $W_w \setminus \mathrm{Min}_{\leqslant'_w}(W)$, provided they do not witness a violation of the sensitivity condition at w for any φ for which we want $\mathcal{N}, w \vDash_h K\varphi$.

Corollary 7.1 (**Soundness and Completeness**)

1. *The system **KT** (equivalently, **ET** plus the RK rule) is sound and complete for C/L-semantics over RA models.*
2. (The Logic of Ranked Relevant Alternatives) *The system **ECNTX** (equivalently, **ET** plus the RAT rule) is sound and complete for D-semantics over* total *RA models.*
3. *The system **ECNT** (equivalently, **ET** plus the RA rule) is sound and complete for D-semantics over RA models.*
4. ***ECNT** is sound (with respect to the full epistemic language) and complete (with respect to the flat fragment) for H/N/S-semantics over CB models.*[50]

The proof of Corollary 7.1 is similar to the alternative completeness proof discussed by van Benthem [8, Section 4.3] for the system **K**.[51]

Proof We only give the proof for part 2, since the proofs for the others are similar. Soundness follows from Theorem 5.2.2. For completeness, we first prove by strong induction on the modal depth $d(\varphi)$ of φ (Def. 2.2) that if φ is D-valid over total RA models, then φ is provable in the system combining **ET** and the RAT rule. If $d(\varphi) = 0$, then the claim is immediate, since our deductive system includes propositional logic. Suppose $d(\varphi) = n + 1$. By the proof of Proposition 5.16, using PL, MP, T, and RE (which is a derived rule given RAT, PL, and MP), we can prove that φ is equivalent to a conjunction φ', each of whose conjuncts is a *T-unpacked* formula (Def. 5.15) of the form

$$\varphi_0 \wedge K\varphi_1 \wedge \cdots \wedge K\varphi_n \to K\psi_1 \vee \cdots \vee K\psi_m. \tag{7.1}$$

The conjunction φ' is valid iff each conjunct of the form of (7.1) is valid. By Theorem 5.2.2, (7.1) is valid iff either condition (a) or condition (c) of Theorem 5.2.2 holds. *Case 1*: (a) holds, so $\varphi_0 \to \bot$ is valid. By the inductive hypothesis, we can derive $\varphi_0 \to \bot$, from which we derive (7.1) using PL and MP. *Case 2*: (c) holds, so for some $\Phi \subseteq \{\varphi_1, \ldots, \varphi_n\}$ and nonempty $\Psi \subseteq \{\psi_1, \ldots, \psi_m\}$,

$$\bigwedge_{\varphi\in\Phi} \varphi \leftrightarrow \bigwedge_{\psi\in\Psi} \psi \tag{7.2}$$

is valid. Since (7.2) is of modal depth less than $n + 1$, by the inductive hypothesis it is provable. From (7.2), we can derive

$$\bigwedge_{\varphi\in\Phi} K\varphi \to \bigvee_{\psi\in\Psi} K\psi \tag{7.3}$$

[50] Corollary 7.1.4 gives an answer, for the flat fragment, to the question posed by van Benthem [8, 153] of what is the epistemic logic of Nozick's notion of knowledge.

[51] The usual canonical model approach used for **K** and other normal modal logics seems more difficult to apply to RA and CB models, since we must use maximally consistent sets of formulas in the epistemic language only (cf. note 61) to guide the construction of both the orderings $\preceq_w$ (resp. $\leqslant_w$) and relation $\rightarrowtriangle$ (resp. D), which must be appropriately related to one another for the truth lemma to hold. In this situation, our alternative approach performs well.

using the RAT rule, from which we can derive (7.1) using PL and MP. Having derived each conjunct of φ' in one of these ways, we can use PL and MP to derive the conjunction itself, which by assumption is provably equivalent to φ.

Next we show by induction on the length of proofs that any proof in the system combining **ET** and RAT can be transformed into an **ECNTX** proof of the same theorem. Suppose that in the first proof, $\varphi_1 \wedge \cdots \wedge \varphi_n \leftrightarrow \psi_1 \wedge \cdots \wedge \psi_m$ has been derived, to which the RAT rule is applied. In the second proof, if $n > 0$, we first derive $K\varphi_1 \wedge \cdots \wedge K\varphi_n \to K(\varphi_1 \wedge \cdots \wedge \varphi_n)$ using C repeatedly (with PL and MP); next, we derive $K(\varphi_1 \wedge \cdots \wedge \varphi_n) \leftrightarrow K(\psi_1 \wedge \cdots \wedge \psi_m)$ by applying the RE rule to $\varphi_1 \wedge \cdots \wedge \varphi_n \leftrightarrow \psi_1 \wedge \cdots \wedge \psi_m$; we then derive $K(\psi_1 \wedge \cdots \wedge \psi_m) \to K\psi_1 \vee \cdots \vee K\psi_m$ using X repeatedly (with PL and MP); finally, we derive $K\varphi_1 \wedge \cdots \wedge K\varphi_n \to K\psi_1 \vee \cdots \vee K\psi_m$ using PL, MP, and earlier steps. If $n = 0$,[52] we first derive $K\top$ using N, then derive $K\top \leftrightarrow K(\psi_1 \wedge \cdots \wedge \psi_m)$ by applying the RE rule to $\top \leftrightarrow \psi_1 \wedge \cdots \wedge \psi_m$, then derive the conclusion of the RAT application using X, PL, and MP. □

For reasons suggested in Section 5.2, I do not consider the systems of Corollary 7.1.2–.4 to be plausible as *epistemic* logics, and therefore I do not consider the basic theories they are based on to be satisfactory theories of knowledge. Nonetheless, we may wish to reason directly about whether one has ruled out the relevant alternatives, whether one's beliefs are sensitive to the truth, etc., and Corollary 7.1 gives principles for these notions. Simply replace the K symbol by a neutral $\Box$ and the newly identified logic **ECNTX**, which I dub *the logic of ranked relevant alternatives*, is of significant independent interest.

With these qualifications in mind, I will make another negative point concerning knowledge. It is easy to derive the K axiom, the star of the epistemic closure debate with its leading role in skeptical arguments, from M, C, RE, and propositional logic. Hence in order to avoid K one must give up one of the latter principles. (For RE, recall that we are considering ideally astute logicians as in Section 2). What is so strange about subjunctivist-flavored theories is that they validate C but not M, which seems to get things backwards. Hawthorne [32, Sections 1.6, 4.6] discusses some of the problems and puzzles, related to the Lottery and Preface Paradoxes (Kyburg [46]; Makinson [54]), to which C leads (also see Goldman [25]). M seems rather harmless by comparison (cf. Williamson [78, Section 12.2]). Interestingly, C also leads to computational difficulties.

Remark 7.2 (*NP vs. PSPACE*) Vardi [72] proved a PSPACE upper bound for the complexity of the system **ECNT**,[53] in agreement with our conclusion in Section 5.3. (Together Corollaries 5.25 and 7.1.2 give a PSPACE upper bound for **ECNTX**.) Vardi also conjectured a PSPACE lower bound for **ECNT**. By contrast, he showed that for any subset of {T, N, M} added to **E**, complexity drops to NP-complete. Hence

[52] If $n = 0$, we can take the left side of the premise/conclusion of RAT to be $\top$, or we can simply take the premise to be $\psi_1 \wedge \cdots \wedge \psi_m$ and the conclusion to be $K\psi_1 \vee \cdots \vee K\psi_m$.

[53] Here I mean either the problem of checking provability/validity or that of checking consistency/satisfiability, given that PSPACE is closed under complementation. When I refer to NP-completeness, I have in mind the consistency/satisfiability problem.

Vardi conjectured that the C axiom is the culprit behind the jump in complexity of epistemic logics from NP to PSPACE.[54] It appears that not only is C more problematic than M epistemologically, but also it makes reasoning about knowledge more computationally costly.[55]

8 Higher-Order Knowledge

In this section, we briefly explore how the theories formalized in Sections 3 and 4 differ with respect to knowledge about one's own knowledge and beliefs. The result is a hierarchical picture (Corollary 8.12) and an open problem for future research. First, we discuss a subtlety concerning higher-order RA knowledge. Second, we relate properties of higher-order subjunctivist knowledge to closure failures.

8.1 Higher-Order Knowledge and Relevant Alternatives

Theorem 5.2 and Corollary 7.1 show that no non-trivial principles of higher order knowledge, such as the controversial 4 axiom $K\varphi \to KK\varphi$ and 5 axiom $\neg K\varphi \to K\neg K\varphi$, are valid over RA models according to either L- or D-semantics. This is so even if we assume that the relation $\rightarrowtriangle$ in our RA models is an equivalence relation (see Remark 5.20), following Lewis [52].

Example 8.1 (*Failure of 4 Axiom*) For the model $\mathcal{M}$ in Fig. 7, in which $\rightarrowtriangle$ is an equivalence relation, observe that $\mathcal{M}, w_1 \nvDash_{l,d} Kp \to KKp$. Since $\text{Min}_{\preceq_{w_1}}(W) \cap \overline{[\![p]\!]} = \{w_2\}$ and $w_1 \not\rightarrowtriangle w_2$ we have $\mathcal{M}, w_1 \vDash_{l,d} Kp$. By contrast, since $w_4 \in \text{Min}_{\preceq_{w_3}}(W) \cap \overline{[\![p]\!]}$ and $w_3 \rightarrowtriangle w_4$, we have $\mathcal{M}, w_3 \nvDash_{l,d} Kp$. It follows that $w_3 \in \text{Min}_{\preceq_{w_1}}(W) \cap \overline{[\![Kp]\!]}$, in which case $\mathcal{M}, w_1 \nvDash_{l,d} KKp$ given $w_1 \rightarrowtriangle w_3$.

According to Williamson [79, 80], "It is not always appreciated that…since Lewis's accessibility relation is an equivalence relation, his account validates not only logical omniscience but the very strong epistemic logic S5" [80, 23n16]. However, Example 8.1 shows that this is not the case if we allow that comparative relevance, like comparative similarity, is possibility-relative, as seems reasonable for a Lewisian

[54] In fact, Allen [2] shows that adding any degree of conjunctive closure, however weak, to the classical modal logic **EMN** results in a jump from NP- to PSPACE-completeness. Adding the full strength of C is sufficient, but not necessary. As far as I know, lower bounds for the complexity of systems with C but without M have not yet been established.

[55] Whether such complexity facts have any philosophical significance seems to be an open question. As a cautionary example, one would not want to argue that it counts in favor of the plausibility of the 5 axiom, $\neg K\varphi \to K\neg K\varphi$, that while the complexity of **K** is PSPACE-complete, for any extension of **K5**, complexity drops to NP-complete [30]. That being said, if we are forced to give up C for epistemological reasons, then its computational costliness in reasoning about knowledge may make us miss it less.

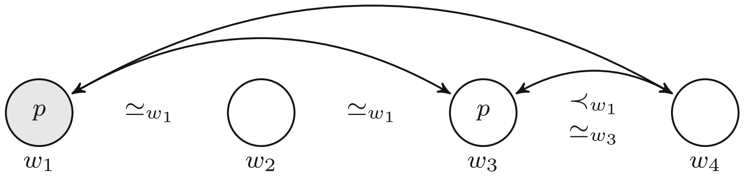

Fig. 7 An RA countermodel for $Kp \to KKp$ in L/D-semantics (partially drawn, reflexive loops omitted)

theory.[56] Other RA theorists are explicit that relevance depends on similarity of worlds (see, e.g., Heller [33, 35]), in which case the former should be world-relative since the latter is. For Williamson's point to hold, we would have to block the likes of Example 8.1 with an additional constraint on our models, such as the following.

Definition 8.2 (Absoluteness) For an RA model $\mathcal{M} = \langle W, \rightarrowtriangle, \preceq, V\rangle$, $\preceq$ is *locally* (resp. *globally*) *absolute* iff for all $w \in W$ and $v \in W_w$ (resp. for all $w, v \in W$), $\preceq_w = \preceq_v$ [49, Section 6.1].

It is noteworthy that absoluteness leads to a collapse of comparative relevance.

Observation 8.3 (Absoluteness and Collapse) Given condition 3b of Definition 3.1, if $\preceq$ is locally absolute, then for all $w \in W$ and $v \in W_w$,

$$\text{Min}_{\preceq_w}(W) = W_w = \text{Min}_{\preceq_v}(W) = W_v.$$

If $\preceq$ is globally absolute, then for all $w \in W$, $\text{Min}_{\preceq_w}(W) = W$.

Lewis [49, 99] rejected absoluteness for comparative similarity because it leads to such a collapse. We note that with the collapse of comparative relevance, the distinction between L- and D-semantics also collapses.

Observation 8.4 (Absoluteness and Collapse cont.) Over locally absolute RA models, L- and D-semantics are equivalent.

The proof of Proposition 8.5, which clarifies the issue raised by Williamson, is essentially the same as that of completeness over standard partition models.

Proposition 8.5 (Completeness of S5) *S5 is sound and complete with respect to L/D-semantics over locally absolute RA models in which $\rightarrowtriangle$ is an equivalence relation.*

[56] It follows from Lewis's [52, 556f] *Rule of Resemblance* that if some $\neg p$-possibility w_2 "saliently resembles" w_1, which is relevant at w_1 by the Rule of Actuality, then w_2 is relevant at w_1, so you must rule out w_2 in order to know p in w_1. Lewis is explicit (555) that by 'actuality' he means the actuality of the subject of knowledge attribution. Hence if we consider your *counterpart* in some w_3, and some $\neg p$-possibility w_4 saliently resembles w_3, then your counterpart must rule out w_4 in order to know p in w_3. However, if salient resemblance is possibility-relative, as comparative similarity is for Lewis, then w_4 may not saliently resemble w_1, in which case you may not need to rule out w_4 in order to know p in w_1. (By Lewis's *Rule of Attention* (559), our attending to w_4 in this way may shift the context $\mathcal{C}$ to a context $\mathcal{C}'$ in which w_4 is relevant, but the foregoing points still apply to $\mathcal{C}$.) This is all that is required for Example 8.1 to be consistent with Lewis's theory.

In general, for locally absolute RA models, the correspondence between properties of $\mathrel{-\!\!\triangleright}$ and modal axioms is exactly as in basic modal logic.

8.2 Higher-Order Knowledge and Subjunctivism

The study of higher-order knowledge becomes more interesting with the subjunctivist theories, especially in connection with our primary concern of closure. According to Nozick [58], the failures of epistemic closure implied by his tracking theory are something that "we must adjust to" (228). This would be easier if problems ended with the closure failures themselves. However, as we will see, the structural features of the subjunctivist theories that lead to these closure failures also lead to problems of higher-order knowledge.

We begin with a definition necessary for stating Fact 8.7 below.

Definition 8.6 (**Outer Necessity**) Let us temporarily extend our language with an *outer necessity* operator $\Box$ [49, Section 1.5] with the truth clause:

$$\mathcal{M}, w \vDash_x \Box\varphi \text{ iff } \forall v \in W_w : \mathcal{M}, v \vDash_x \varphi.$$

We call the language with K, B, and $\Box$ the *epistemic-doxastic-alethic* language. Define the possibility operator by $\Diamond\varphi := \neg\Box\neg\varphi$, and let $\hat{K}\varphi := \neg K\neg\varphi$.

Fact 8.7 below shows that if *sensitivity* (Def. 4.3) is necessary for knowledge, and if there is any counterfactually accessible world in which an agent believes φ but φ is false, then the agent cannot know that her belief that φ is not false—*even if she knows that φ is true*.[57] The proof appears in many places [17, 44, 66, 67, 73, 75].

Fact 8.7 (**Possibility and Sensitivity**) $\Diamond(B\varphi \wedge \neg\varphi) \rightarrow \hat{K}(B\varphi \wedge \neg\varphi)$ *is H/N-valid, but not S-valid.*

Since $Kp \wedge \Diamond(Bp \wedge \neg p)$ is satisfiable, $Kp \rightarrow K\neg(Bp \wedge \neg p)$ is not H/N-valid by Fact 8.7, so $Kp \rightarrow K(\neg Bp \vee p)$ is not H/N-valid. Hence Fact 8.7 is related to the failure of closure under disjunctive addition. Clearly $\Diamond\psi \rightarrow \hat{K}\psi$ is not H/N-valid for all ψ. Related to Fact 8.7, Fact 8.8 (used for Corollary 8.12) shows that limited forms of closure, including closure under disjunctive addition, hold when higher-order knowledge of $B\varphi \rightarrow \varphi$ or $\hat{K}\varphi \rightarrow \varphi$ is involved.

Fact 8.8 (**Higher-Order Closure**)

1. $K(B\varphi \rightarrow \varphi) \rightarrow K((B\varphi \rightarrow \varphi) \vee \psi)$ *is H/S-valid, but not N-valid;*
2. $B\varphi \wedge K(B\varphi \rightarrow \varphi) \rightarrow K\varphi$ *is H/S-valid, but not N-valid;*
3. $B\varphi \wedge K(\hat{K}\varphi \rightarrow \varphi) \rightarrow K\varphi$ *is H/S-valid, but not N-valid;*
4. $K(\varphi \wedge \psi) \wedge K(\hat{K}\varphi \rightarrow \varphi) \rightarrow K\varphi$ *is H/S-valid, but not D/N-valid;*

[57] More precisely, she cannot know that she does not have a false belief that φ [6]. As Becker [6] in effect proves, $BB\varphi \wedge K\varphi \rightarrow K(B\varphi \wedge \varphi)$ is H-valid (and hence S-valid).

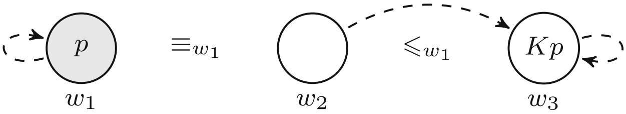

Fig. 8 A CB model satisfying $K(p \wedge \neg Kp)$ in H/N/S-semantics (partially drawn)

5. $K\varphi \wedge K\psi \wedge K(\hat{K}(\varphi \vee \psi) \to (\varphi \vee \psi)) \to K(\varphi \vee \psi)$ *is H/N/S-valid, but not D-valid (over total RA models).*

While some consider Fact 8.7 to be a serious problem for sensitivity theories, Fact 8.9 seems even worse for subjunctivist-flavored theories in general: according to the ones we have studied, it is possible for an agent to know the classic example of an unknowable sentence, $p \wedge \neg Kp$ [23]. Williamson [78, 279] observes that $p \wedge \neg Kp$ is knowable according to the sensitivity theory. We observe that it is also knowable according to the safety theory.[58]

Fact 8.9 (**Moore-Fitch Sentences**) *$K(p \wedge \neg Kp)$ is satisfiable in RA models under D-semantics and in CB models under H/N/S-semantics.*

Proof It is immediate from Theorem 5.2 that $\neg K(p \wedge \neg Kp)$ is not D-valid.[59]

We give a simple satisfying CB model $\mathcal{M}$ for H/N/S-semantics in Fig. 8. Assume that $\leqslant_{w_3}$ is any appropriate preorder such that $\mathcal{M}, w_3 \vDash_{h,n,s} Kp$. It will not matter whether $w_1 \equiv_{w_1} w_2 \equiv_{w_1} w_3$ or $w_1 \equiv_{w_1} w_2 <_{w_1} w_3$.

Given $w_2 \in \text{Min}_{\leqslant_{w_1}}(W)$ and $\mathcal{M}, w_2 \vDash \neg p \wedge Bp$, the safety condition for Kp fails at w_1, so $\mathcal{M}, w_1 \vDash_s p \wedge \neg Kp$. Then since $D(w_1) = \{w_1\}$ (recall Notation 5.26), $\mathcal{M}, w_1 \vDash_s B(p \wedge \neg Kp)$, so the belief condition for $K(p \wedge \neg Kp)$ holds at w_1. For $i \geq 2$, given $\mathcal{M}, w_i \vDash BKp$, we have $\mathcal{M}, w_i \nvDash B(p \wedge \neg Kp)$. It follows that for all $v \in \text{Min}_{\leqslant_{w_1}}(W)$, $\mathcal{M}, v \vDash_s B(p \wedge \neg Kp) \to p \wedge \neg Kp$. Hence the safety condition for $K(p \wedge \neg Kp)$ holds at w_1, so $\mathcal{M}, w_1 \vDash_s K(p \wedge \neg Kp)$. One can check that $\mathcal{M}, w_1 \vDash_{h,n} K(p \wedge \neg Kp)$ as well. For H/N-semantics, the model $\mathcal{N}$ in Fig. 9, which has the same basic structure as Williamson's [78, 279] example, also satisfies $K(p \wedge \neg Kp)$ at w_1. Assume $\leqslant_{w_2}$ is any appropriate preorder such that $\mathcal{N}, w_2 \vDash_{h,n} Kp$.[60] (Whether $w_1 \equiv_{w_1} w_2$ or $w_1 <_{w_1} w_2$ does not matter). □

It is not difficult to tell a story with the structure of Fig. 8, illustrating that the safety theory allows $K(p \wedge \neg Kp)$, just as Williamson tells a story with the structure of Fig. 9, illustrating that the tracking theory allows $K(p \wedge \neg Kp)$.

[58] One difference between Fact 8.7 and Fact 8.9 is that the former applies to any theory for which sensitivity is a necessary condition for knowledge, whereas the latter could in principle be blocked by theories that propose other necessary conditions for knowledge in addition to sensitivity or safety. What Fact 8.9 shows is that sensitivity and safety theorists have some explaining to do about what they expect to block such a counterintuitive result.

[59] Rewrite $\neg K(p \wedge \neg Kp)$ as $K(p \wedge \neg Kp) \to \bot$. T-unpacking gives $p \wedge \neg Kp \wedge K(p \wedge \neg Kp) \to \bot$ and then $p \wedge K(p \wedge \neg Kp) \to Kp$, which fails (a), (c), and (d) of Theorem 5.2.

[60] One can of course add more worlds to W_{w_2} than are shown in Fig. 9.

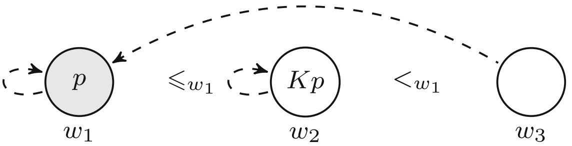

Fig. 9 A CB model satisfying $K(p \wedge \neg Kp)$ in H/N-semantics (partially drawn)

Fact 8.9 is related to the fact that closure under conjunction elimination is not valid. Otherwise $K(p \wedge \neg Kp)$ would be unsatisfiable; for by veridicality, $K(p \wedge \neg Kp) \to \neg Kp$ is valid, and given closure under conjunction elimination, $K(p \wedge \neg Kp) \to Kp$ would also be valid. However, Fact 8.10 shows that K does partially distribute over conjunctions of special forms in S-semantics.

Fact 8.10 (Higher-Order Closure cont.)

1. *$K(\varphi \wedge \neg K\varphi) \to K\neg K\varphi$ is S-valid, but not D/H/N-valid.*
2. *$K((\varphi \vee \psi) \wedge (B\varphi \to \varphi)) \to K(B\varphi \to \varphi)$ is S-valid, but not H/N-valid.*

What Facts 8.9 and 8.7 show is that in order to fully calculate the costs of closure failures, one must take into account their ramifications in the realm of higher-order knowledge. Combining Facts 8.7, 8.8, and 8.10 with results from earlier sections, we arrive at a picture of the relations between the sets of valid principles according to D-, H-, N-, and S-semantics, respectively, given by Corollary 8.12 below.[61] First we need the following definition.

Definition 8.11 (Theories and Model Classes) For a class S of models, let $\mathrm{Th}^x_{\mathcal{L}}(\mathsf{S})$ be the set of formulas in the language $\mathcal{L}$ that are valid over S according to X-semantics. Let RAT be the class of all total RA models, RA the class of all RA models, and CB the class of all CB models.

Corollary 8.12 (Hierarchies)

1. *For the flat fragment $\mathcal{L}_f$ of the epistemic language,*

$$Th^n_{\mathcal{L}_f}(\mathsf{CB}) = Th^h_{\mathcal{L}_f}(\mathsf{CB}) = Th^s_{\mathcal{L}_f}(\mathsf{CB}) = Th^d_{\mathcal{L}_f}(\mathsf{RA}) \subsetneq Th^d_{\mathcal{L}_f}(\mathsf{RAT}).$$

2. *For the epistemic language $\mathcal{L}_e$,*

$$Th^d_{\mathcal{L}_e}(\mathsf{RA}) \subsetneq Th^n_{\mathcal{L}_e}(\mathsf{CB}) \subsetneq Th^h_{\mathcal{L}_e}(\mathsf{CB}) \subsetneq Th^s_{\mathcal{L}_e}(\mathsf{CB});$$

$$Th^d_{\mathcal{L}_e}(\mathsf{RA}) \subsetneq Th^d_{\mathcal{L}_e}(\mathsf{RAT}) \not\subseteq Th^s_{\mathcal{L}_e}(\mathsf{CB}); \;\; Th^n_{\mathcal{L}_e}(\mathsf{CB}) \not\subseteq Th^d_{\mathcal{L}_e}(\mathsf{RAT}).$$

[61] If we require more properties of the D relation, then more principles will be valid in H/N/S-semantics—obviously for the B operator, but also for the interaction between K and B. For example, if require that D be *dense*, so $BB\varphi \to B\varphi$ is valid, then $BB\varphi \to KB\varphi$ is H/S-valid. If we also require that D be transitive, so $B\varphi \to BB\varphi$ is valid, then $B\varphi \to KB\varphi$ is H/N/S-valid. As Kripke [44, 183] in effect observes, if $B\varphi \leftrightarrow BB\varphi$ is valid, then (for propositional φ) $\mathcal{M}, w \vDash_h K\varphi$ implies $\mathcal{M}, w \vDash_n K(\varphi \wedge B\varphi)$, so whenever $\mathcal{M}, w \vDash_h K\varphi$ but $\mathcal{M}, w \nvDash_n K\varphi$ (because adherence is not satisfied), $K(\varphi \wedge B\varphi) \to K\varphi$ fails according to N-semantics, an extreme closure failure.

3. *For the epistemic-doxastic language $\mathcal{L}_d$,*

$$Th^n_{\mathcal{L}_d}(\mathsf{CB}) \subsetneq Th^h_{\mathcal{L}_d}(\mathsf{CB}) \subsetneq Th^s_{\mathcal{L}_d}(\mathsf{CB}).$$

4. *For the epistemic-doxastic-alethic language $\mathcal{L}_a$,*

$$Th^n_{\mathcal{L}_a}(\mathsf{CB}) \subsetneq Th^h_{\mathcal{L}_a}(\mathsf{CB});\; Th^n_{\mathcal{L}_a}(\mathsf{CB}) \nsubseteq Th^s_{\mathcal{L}_a}(\mathsf{CB}) \nsubseteq Th^h_{\mathcal{L}_a}(\mathsf{CB}).$$

Proof Part 1 follows from Corollary 7.1 and Fact 5.7. Part 2 follows from Corollary 7.1, Propositions 6.2–6.3, and Facts 8.8.5, 8.8.4, 8.10.1, and 5.7. Part 3 follows from Propositions 6.2–6.3 and Facts 8.8 and 8.10. Part 4 follows from Proposition 6.2 (which clearly extends to $\mathcal{L}_a$) and Facts 8.8, 8.7, and 8.10. □

In this section we have focused on the implications of D/H/N/S-semantics for higher-order *knowledge*, especially in connection with epistemic closure. However, if we take the point of view suggested earlier (Sections 1, 5 and 7), according to which our results can be interpreted as results about desirable epistemic properties other than knowledge, then exploring higher-order phenomena in D/H/N/S-semantics is part of understanding these other properties. Along these lines, we conclude this section with an open problem for future research.

Problem 8.13 (**Axiomatization**) Axiomatize the theory of counterfactual belief models according to H-, N-, or S-semantics for the full epistemic, epistemic-doxastic, or epistemic-doxastic-alethic language.[62]

[62] If we extend the language of Definition 2.2 so that we can describe different parts of our CB models independently, e.g., by adding the belief operator B for the doxastic relation D or a counterfactual conditional $\mathrel{\Box\!\!\rightarrow}$ for the similarity relations $\leqslant_w$, then the problem of axiomatization becomes easier. For S-semantics, which does not use the structure of any $\leqslant_w$ relation beyond $\mathrm{Min}_{\leqslant_w}(W)$, just adding B to the language makes the axiomatization problem easy. As one can prove by a standard canonical model construction, for completeness it suffices to combine the logic **KD** for B with the axiom $K\varphi \to B\varphi$ and the rule

$$\mathrm{SA}.\ \frac{(B\varphi_1 \to \varphi_1)\wedge\cdots\wedge(B\varphi_n \to \varphi_n) \to (B\psi \to \psi)}{K\varphi_1 \wedge\cdots\wedge K\varphi_n \to (B\psi \to K\psi)}\ {\scriptstyle (n\geq 0)}.$$

For H/N-semantics, adding not only B but also a counterfactual $\mathrel{\Box\!\!\rightarrow}$ (with the Lewisian semantics outlined in Section 4) makes the axiomatization problem easy. For example, for N-semantics we can combine **KD** for B with a complete system for counterfactuals (no interaction axioms between B and $\mathrel{\Box\!\!\rightarrow}$ are needed), plus $K\varphi \to B\varphi$ and $K\varphi \leftrightarrow B\varphi \wedge (\neg\varphi \mathrel{\Box\!\!\rightarrow} \neg B\varphi) \wedge (B\varphi \mathrel{\Box\!\!\rightarrow} \varphi)$. The problem with obtaining easy axiomatizations by extending the language in this way is that the resulting systems give us little additional insight. The interesting properties of knowledge are hidden in the axioms that combine several operators, each with different properties. Although in a complete system for the extended language we can of course derive all principles that could appear in any sound system for a restricted language, this fact does not tell us what those principle are or which set of them is complete with respect to the restricted language. Corollary 7.1 and Facts 8.7, 8.8, and 8.10 suggest that more illuminating principles may appear as axioms if we axiomatize the S-theory of CB models in the epistemic language or the H/N-theory of CB models in the epistemic-doxastic(-alethic) language.

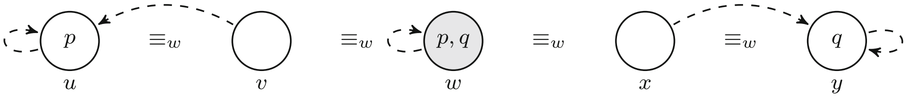

Fig. 10 A CB countermodel for $K(p \wedge q) \rightarrow Kp \vee Kq$ in H/N/S-semantics (partially drawn)

9 Theory Parameters and Closure

In this section, we return to the issue raised in Section 5.5 about the sources of closure failure. Analysis of Theorem 5.2 shows that two parameters of a modal theory of knowledge affect whether closure holds. In Section 3, we identified one: the ∀∃ vs. ∃∀ choice of the relevancy set. Both L- and S-semantics have an ∃∀ setting of this parameter (recall Observation 4.5). However, closure holds in L-semantics but fails in S-semantics. The reason for this is the second theory parameter: the notion of ruling out. With the Lewis-style notion of ruling out in L/D-semantics, a world v is either ruled out at w or not. By contrast, with the notions of ruling out implicit in S/H/N-semantics, we cannot say independently of a proposition in question whether v is ruled out at w.

For example, in the CB model in Fig. 10, v is among the closest worlds to the actual world w. We may say that v is ruled out *as an alternative for* $p \wedge q$, in the sense that while $p \wedge q$ is false at v, the agent does not believe $p \wedge q$ at v (but rather $p \wedge \neg q$). However, v is not ruled out *as an alternative for* p, for p is false at v and yet the agent believes p at v. This explains the consequence of Theorem 5.2 that $K(p \wedge q) \rightarrow Kp$ is not valid in S-semantics, because one may *safely* believe $p \wedge q$ at a world w even though one does not safely believe p at w. Note that the example also applies to sensitivity theories, for which we can again only say whether v is ruled out *as an alternative for* a given φ.

The distinction between the two notions of ruling out (RO) is again that of ∀∃ vs. ∃∀, as in the case of $\mathsf{RS}_{\forall\exists}$ vs. $\mathsf{RS}_{\exists\forall}$ in Section 3. Let us state the distinction in terms of possibilities that are *not* ruled out, possibilities that are *uneliminated*:

> According to an $\mathsf{RO}_{\forall\exists}$ theory, for every context $\mathcal{C}$, world w, and (∀) proposition P, there is (∃) a set of worlds $\mathsf{u}_{\mathcal{C}}(P, w) \subseteq \overline{P}$ *uneliminated at w as alternatives for P*, such that if any world in $\mathsf{u}_{\mathcal{C}}(P, w)$ is relevant (i.e., in $\mathsf{r}_{\mathcal{C}}(P, w)$), then the agent does not know P in w (relative to $\mathcal{C}$).
>
> According to an $\mathsf{RO}_{\exists\forall}$ theory, for every context $\mathcal{C}$ and world w, there is (∃) a set of worlds $\mathsf{U}_{\mathcal{C}}(w)$ *uneliminated at* w, such that for every (∀) proposition P, if any world in $\mathsf{U}_{\mathcal{C}}(w) \cap \overline{P}$ is relevant (i.e., in $\mathsf{r}_{\mathcal{C}}(P, w)$), then the agent does not know P in w (relative to $\mathcal{C}$).

Every $\mathsf{RO}_{\exists\forall}$ theory is a $\mathsf{RO}_{\forall\exists}$ theory (with $\mathsf{u}_{\mathcal{C}}(P, w) = \mathsf{U}_{\mathcal{C}}(w) \cap \overline{P}$), but when I refer to $\mathsf{RO}_{\forall\exists}$ theories I have in mind those that are not $\mathsf{RO}_{\exists\forall}$. As noted, L/D-semantics formalize $\mathsf{RO}_{\exists\forall}$ theories, with $\twoheadrightarrow(w)$ (Notation 5.12) in the role of $\mathsf{U}(w)$, while S/H/N-semantics formalize $\mathsf{RO}_{\forall\exists}$ theories, given the role of belief in their notions of ruling out, noted above (see [38, Section 3.3.2]).

Table 2 Parameter settings and closure failures

Theory	Formalization	Relevancy set	Ruling out	Closure failures
RA	L-semantics	∃∀	∃∀	None
RA	D-semantics	∀∃	∃∀	Theorem 5.2
Safety	S-semantics	∃∀	∀∃	Theorem 5.2
Tracking	H/N-semantics	∀∃	∀∃	Theorem 5.2

Consider the parallel between $\mathsf{RO}_{\forall\exists}$ and $\mathsf{RS}_{\forall\exists}$ parameter settings: given a ∀∃ setting of the RO (resp. RS) parameter, a $(\neg\varphi \wedge \neg\psi)$-world that is ruled out as an alternative for φ (resp. that must be ruled out in order to know φ) may not be ruled out as an alternative for ψ (resp. may not be such that it must be ruled out in order to know ψ), because whether the world is ruled out or not (resp. relevant or not) depends on the proposition in question, as indicated by the ∀ *propositions* ∃ *set of uneliminated (resp. relevant) worlds* quantifier order. As the example of Fig. 10 shows, the $\mathsf{RO}_{\forall\exists}$ setting for safety explains why closure fails in S-semantics, despite its $\mathsf{RS}_{\exists\forall}$ setting.

Table 2 summarizes the relation between the two theory parameters and closure failures. Not all theories with $\mathsf{RS}_{\forall\exists}$ or $\mathsf{RO}_{\forall\exists}$ settings must have the *same* closure failures as described by Theorem 5.2. Elsewhere I show that as a result of their particular $\mathsf{RO}_{\forall\exists}$ character, variants of subjunctivism, such as DeRose [17] modified tracking theory and the safety theory with *bases*, do not avoid serious closure failures [38, Sections 2.10.1, 2.D]. However, in part II we will see how a kind of generalized $\mathsf{RS}_{\forall\exists}$ theory can avoid the worst of the subjunctivist-flavored theories, while still stopping short of full closure. This theory will solve the Dretskean closure dilemma raised at the end of Section 3.

10 Conclusion of Part I

In this paper, we have investigated an area where epistemology and epistemic logic naturally meet: the debate over epistemic closure, involving two of the most influential views in contemporary epistemology—relevant alternatives and subjunctivism. Our model-theoretic approach helped to illuminate the structural features of RA and subjunctivist theories that lead to closure failure, as well as the precise extent of their closure failures in Theorem 5.2.

When understood as theories of *knowledge*, the basic subjunctivist-flavored theories formalized by D/H/N/S-semantics have a bad balance of closure properties. Not only do they invalidate very plausible closure principles (recall Section 5), illustrating the problem of *containment* (recall Section 1), but also they validate some questionable ones (recall Section 7). The theories formalized by C- and L-semantics also have their problems. On the one hand, the idea that knowledge requires ruling out *all* possibilities of error, reflected in C-semantics, makes knowing too hard, giving us the problem of *skepticism* (recall Sections 2 and 3). On the other hand, the idea that knowledge of contingent empirical truths can be acquired with *no* requirement of eliminating possibilities, reflected in L-semantics (and S-semantics), seems to make knowing too easy, giving us the problem of *vacuous knowledge* (recall Sections 3 and 4). An attraction of D/H/N-semantics is that they avoid these problems. But they do so at a high cost when it comes to closure.

In Part II, I will propose a new picture of knowledge that avoids the problems of skepticism and vacuous knowledge, without the high-cost closure failures of the subjunctivist-flavored theories. As we shall see, the model-theoretic epistemic-logical approach followed here can help us not only to better understand epistemological problems, but also to discover possible solutions.

The results of this paper motivate some methodological reflections on our approach. In epistemology, a key method of theory assessment involves considering the verdicts issued by different theories about which knowledge claims are true in a particular scenario. This is akin to considering the verdicts issued by different semantics about which epistemic formulas are true in a particular model. All of the semantics we studied can issue different verdicts for the same model. Moreover, theorists who favor different theories/semantics may represent a scenario with different models in the first place. Despite these differences, there are systematic relations between the RA, tracking, and safety perspectives represented by our semantics. In several cases, we have seen that any model viewed from one perspective can be transformed into a model that has an equivalent epistemic description from a different perspective (Propositions 6.1–6.3). As we have also seen, when we rise to the level of truth in *all models*, of validity, differences may wash away, revealing unity on a higher level. Theorem 5.2 provided such a view, showing that four different epistemological pictures validate essentially the same epistemic closure principles. Against this background of similarity, subtle differences within the RA/subjunctivist family appear more clearly. The picture offered by total relevant alternatives models lead to a *logic of ranked relevant alternatives*, interestingly different from the others (Corollary 7.1). In the realm of higher-order knowledge, there emerged hierarchies in the strength of different theories (Corollary 8.12).

For some philosophers, a source of hesitation about epistemic logic is the degree of idealization. In basic systems of epistemic logic, agents know all the logical consequences of what they know, raising the "problem of logical omniscience" noted in Section 1. However, in our setting, logical omniscience is a feature, not a bug. Although in our formalizations of the RA and subjunctivist theories, agents do not know all the logical consequences of what they know, due to failures of epistemic closure, they are still logically omniscient in another sense. For as "ideally astute logicians" (recall Section 2), they know all logically valid principles, and they believe all the logical consequences of what they believe. These assumptions allow us to distinguish failures of epistemic closure that are due to fact that finite agents do not always "put two and two together" from failures of epistemic closure that are due to the special conditions on knowledge posited by the RA and subjunctivist theories.[63] This shows the positive role that idealization can play in epistemology, as it does in science.

Acknowledgments I wish to thank everyone who provided helpful feedback on earlier versions of this paper, especially Johan van Benthem and Krista Lawlor, as well as Davide Grossi, Peter Hawke, Tomohiro Hoshi, Thomas Icard, Alistair Isaac, Neil Van Leeuwen, Helen Longino, Eric Pacuit, Sherrilyn Roush, and an anonymous referee for JPL.

References

1. Adams, F., Barker, J.A., Figurelli, J. (2012). Towards closure on closure. *Synthese*, *188*(2), 179–196.
2. Allen, M. (2005). Complexity results for logics of local reasoning and inconsistent belief. In R. van der Meyden (Ed.), *Proceedings of the tenth conference on theoretical aspects of rationality and knowledge (TARK X)* (pp. 92–108). National University of Singapore.
3. Alspector-Kelly, M. (2011). Why safety doesn't save closure. *Synthese*, *183*(2), 127–142.
4. Audi, R. (1988). *Belief, justification, and knowledge*. Belmont: Wadsworth Publishing.
5. Austin, J. (1946). Other minds. *Proceedings of the Aristotelian Society*, *20*, 148–187.
6. Becker, K. (2006). Is counterfactual reliabilism compatible with higher-level knowledge? *Dialectica*, *60*(1), 79–84.
7. Becker, K. (2007). *Epistemology modalized*. New York: Routledge.
8. van Benthem, J. (2010). *Modal logic for open minds*. Stanford: CSLI Publications.
9. Black, T. (2010). Modal and anti-luck epistemology. In S. Bernecker & D. Pritchard (Eds.), *The routledge companion to epistemology* (pp. 187–198). New York: Routledge.
10. Brueckner, A. (2004). Strategies for refuting closure for knowledge. *Analysis*, *64*(4), 333–335.
11. Chalmers, D.J. (2011). The nature of epistemic space. In A. Egan & B. Weatherson (Eds.), *Modality, epistemic* (pp. 60–107). New York: Oxford University Press.
12. Chellas, B.F. (1980). *Modal logic: an introduction*. New York: Cambridge University Press.
13. Cohen, S. (1988). How to be a fallibilist. *Philosophical Perspectives*, *2*, 91–123.
14. Cohen, S. (2002). Basic knowledge and the problem of easy knowledge. *Philosophy and Phenomenological Research*, *65*(2), 309–329.
15. Comesaña, J. (2007). Knowledge and subjunctive conditionals. *Philosophy Compass*, *2*(6), 781–791.
16. Cross, C.B. (2008). Antecedent-relative comparative world similarity. *Journal of Philosophical Logic*, *37*(2), 101–120.
17. DeRose, K. (1995). Solving the skeptical problem. *The Philosophical Review*, *104*(1), 1–52.
18. DeRose, K. (2009). *The case for contextualism: knowledge, skepticism and context*. New York: Oxford University Press.
19. Dretske, F. (1970). Epistemic operators. *The Journal of Philosophy*, *67*(24), 1007–1023.
20. Dretske, F. (1971). Conclusive reasons. *Australasian Journal of Philosophy*, *49*(1), 1–22.
21. Dretske, F. (1981). The pragmatic dimension of knowledge. *Philosophical Studies*, *40*(3), 363–378.
22. Dretske, F. (2005). The case against closure. In M. Steup & E. Sosa (Eds.), *Contemporary debates in epistemology* (pp. 13–26). Malden: Blackwell Publishing.

[63] Recall note 28. Williamson [81, 256] makes a similar point, namely that it can be useful to assume logical omniscience in order to discern the specific epistemic effects of limited powers of perceptual discrimination, as opposed to limited logical powers.

23. Fitch, F.B. (1963). A logical analysis of some value concepts. *The Journal of Symbolic Logic*, *28*(2), 135–142.
24. Friedman, N., & Halpern, J.Y. (1994). On the complexity of conditional logics. In J. Doyle, E. Sandwell, P. Torasso (Eds.), *Proceedings of the fourth international conference on principles of knowledge representation and reasoning (KR'94)* (pp. 202–213). San Francisco: Morgan Kaufman.
25. Goldman, A.H. (1975). A note on the conjunctivity of knowledge. *Analysis*, *36*(1), 5–9.
26. Goldman, A.I. (1976). Discrimination and perceptual knowledge. *The Journal of Philosophy*, *73*(20), 771–791.
27. Goldman, A.I. (1986). *Epistemology and cognition*. Cambridge: Harvard University Press.
28. Halpern, J.Y. (1995). The effect of bounding the number of primitive propositions and the depth of nesting on the complexity of modal logic. *Artificial Intelligence*, *75*(2), 361–372.
29. Halpern, J.Y., & Pucella, R. (2011). Dealing with logical omniscience: expressiveness and pragmatics. *Artificial Intelligence*, *175*(1), 220–235.
30. Halpern, J.Y., & Rêgo, L.C. (2007). Characterizing the NP-PSPACE gap in the satisfiability problem for modal logic. *Journal of Logic and Computation*, *17*(4), 795–806.
31. Harman, G., & Sherman, B. (2004). Knowledge, assumptions, lotteries. *Philosophical Issues*, *14*(1), 492–500.
32. Hawthorne, J. (2004). *Knowledge and lotteries*. New York: Oxford University Press.
33. Heller, M. (1989). Relevant alternatives. *Philosophical Studies*, *55*(1), 23–40.
34. Heller, M. (1999a). Relevant alternatives and closure. *Australasian Journal of Philosophy*, *77*(2), 196–208.
35. Heller, M. (1999b). The proper role for contextualism in an anti-luck epistemology. *Noûs*, *33*(s13), 115–129.
36. Hintikka, J. (1962). *Knowledge and belief: an introduction to the logic of the two notions*. Ithaca: Cornell University Press.
37. Holliday, W.H. (2012a). Epistemic logic, relevant alternatives, and the dynamics of context. In D. Lassiter & M. Slavkovik (Eds.), *New directions in logic, language and computation, lecture notes in computer science* (Vol. 7415, pp. 109–129).
38. Holliday, W.H. (2012b). *Knowing what follows: epistemic closure and epistemic logic*. PhD thesis, Stanford University, revised version, ILLC dissertation series DS-2012-09.
39. Holliday, W.H. (2013a). Epistemic logic and epistemology. In S.O. Hansson & V.F. Hendricks (Eds.), *Handbook of formal philosophy*. Dordrecht: Springer, forthcoming.
40. Holliday, W.H. (2013b). Fallibilism and multiple paths to knowledge. In T.S. Gendler & J. Hawthorne (Eds.), *Oxford studies in epistemology* (Vol. 5). New York: Oxford University Press, forthcoming.
41. Holliday, W.H., & Perry, J. (2013c). Roles, rigidity, and quantification in epistemic logic. In A. Baltag & S. Smets (Eds.), *Johan F. A. K. van Benthem on logical and informational dynamics*. Dordrecht: Springer, forthcoming.
42. King, J.C. (2007). What in the world are ways things might have been? *Philosophical Studies*, *133*(3), 443–453.
43. Kripke, S. (1963). Semantical considerations on modal logic. *Acta Philosophica Fennica*, *16*, 83–94.
44. Kripke, S. (2011). Nozick on knowledge. In *Philosophical troubles: collected papers* (Vol. 1, pp. 162–224). New York: Oxford University Press.
45. Kvanvig, J.L. (2006). Closure principles. *Philosophy Compass*, *1*(3), 256–267.
46. Kyburg, H.E. Jr. (1961). *Probability and the logic of rational belief*. Middletown: Wesleyan University Press.
47. Lawlor, K. (2005). Living without closure. *Grazer Philosophische Studien*, *69*, 25–49.
48. Lewis, D. (1971). Completeness and decidability of three logics of counterfactual conditionals. *Theoria*, *37*(1), 74–85.
49. Lewis, D. (1973). *Counterfactuals*. Oxford: Basil Blackwell.
50. Lewis, D. (1981). Ordering semantics and premise semantics for counterfactuals. *Journal of Philosophical Logic*, *10*(2), 217–234.
51. Lewis, D. (1986). *On the plurality of worlds*. Oxford: Basil Blackwell.
52. Lewis, D. (1996). Elusive knowledge. *Australasian Journal of Philosophy*, *74*(4), 549–567.
53. Luper-Foy, S. (1984). The epistemic predicament: knowledge, Nozickian tracking, and scepticism. *Australasian Journal of Philosophy*, *62*(1), 26–49.
54. Makinson, D. (1965). The paradox of the preface. *Analysis*, *25*(6), 205–207.
55. McGinn, C. (1984). The concept of knowledge. *Midwest Studies in Philosophy*, *9*(1), 529–554.

56. Murphy, P. (2005). Closure failures for safety. *Philosophia*, *33*, 331–334.
57. Murphy, P. (2006). A strategy for assessing closure. *Erkenntnis*, *65*(3), 365–383.
58. Nozick, R. (1981). *Philosophical explanations*. Cambridge: Harvard University Press.
59. Papadimitriou, C.H. (1994). *Computational complexity*. Reading: Addison-Wesley.
60. Pritchard, D. (2005). *Epistemic luck*. New York: Oxford University Press.
61. Pritchard, D. (2008). Sensitivity, safety, and antiluck epistemology. In J. Greco (Ed.), *The Oxford handbook of skepticism* (pp. 437–455). New York: Oxford University.
62. Roush, S. (2005). *Tracking truth: knowledge, evidence and science*. New York: Oxford University Press.
63. Roush, S. (2012). Sensitivity and closure. In K. Becker & T. Black (Eds.), *The sensitivity principle in epistemology* (pp. 242–268). New York: Cambridge University Press.
64. Rysiew, P. (2006). Motivating the relevant alternatives approach. *Canadian Journal of Philosophy*, *36*(2), 259–279.
65. Sherman, B., & Harman, G. (2011). Knowledge and assumptions. *Philosophical Studies*, *156*(1), 131–140.
66. Sosa, E. (1996). Postscript to proper functionalism and virtue epistemology. In J. Kvanvig (Ed.), *Warrant in contemporary epistemology* (pp. 271–281). Totowa: Rowman and Littlefield.
67. Sosa, E. (1999). How to defeat opposition to Moore. *Noûs*, *33*(13), 141–153.
68. Stalnaker, R. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory* (Vol. 2, pp. 98–112). Oxford: Basil Blackwell.
69. Stalnaker, R. (1991). The problem of logical omniscience, I. *Synthese*, *89*(3), 425–440.
70. Stine, G. (1976). Skepticism, relevant alternatives, and deductive closure. *Philosophical Studies*, *29*(4), 249–261.
71. Stroud, B. (1984). *The significance of philosophical scepticism*. New York: Oxford University Press.
72. Vardi, M.Y. (1989). On the complexity of epistemic reasoning. In *Proceedings of the 4th IEEE symposium on logic in computer science* (pp. 43–252).
73. Vogel, J. (1987). Tracking, closure, and inductive knowledge. In S. Luper-Foy (Ed.), *The possibility of knowledge: nozick and his critics* (pp. 197–215). Totowa: Rowman and Littlefield.
74. Vogel, J. (1999). The new relevant alternatives theory. *Noûs*, *33*(s13), 155–180.
75. Vogel, J. (2000). Reliabilism leveled. *The Journal of Philosophy*, *97*(11), 602–623.
76. Vogel, J. (2007). Subjunctivitis. *Philosophical Studies*, *134*(1), 73–88.
77. Warfield, T.A. (2004). When epistemic closure does and does not fail: a lesson from the history of epistemology. *Analysis*, *64*(1), 35–41.
78. Williamson, T. (2000). *Knowledge and its limits*. New York: Oxford University Press.
79. Williamson, T. (2001). Comments on Michael Williams' contextualism, externalism, and epistemic standards. *Philosophical Studies*, *103*(1), 25–33.
80. Williamson, T. (2009). Probability and danger. *The Amherst Lecture in Philosophy*, *4*, 1–35. http://www.amherstlecture.org/williamson2009/.
81. Williamson, T. (2010). Interview. In V. Hendricks & O. Roy (Eds.), *Epistemic logic: 5 questions* (pp. 249–261). New York: Automatic.