Epistemic Logic, Relevant Alternatives, and the Dynamics of Context*

Wesley H. Holliday

Abstract. According to the Relevant Alternatives (RA) Theory of knowledge, knowing that something is the case involves *ruling out* (*only*) *the relevant alternatives*. The conception of knowledge in epistemic logic also involves the elimination of possibilities, but without an explicit distinction, among the possibilities consistent with an agent's information, between those *relevant* possibilities that an agent must rule out in order to know and those remote, far-fetched or otherwise irrelevant possibilities. In this article, I propose formalizations of two versions of the RA theory. Doing so clarifies a famous debate in epistemology, pitting Fred Dretske against David Lewis, about whether the RA theorist should accept the principle that *knowledge is closed under known implication*, familiar as the K axiom in epistemic logic. Dretske's case against closure under known implication leads to a study of other closure principles, while Lewis's defense of closure by appeal to the claimed *context sensitivity* of knowledge attributions leads to a study of the dynamics of context. Having followed the first lead at length in other work, here I focus more on the second, especially on logical issues associated with developing a dynamic epistemic logic of context change over models for the RA theory.

1 Introduction

Example 1 (Medical Diagnosis). Suppose that 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` might still have made the diagnosis of c, since the physical exam would not have revealed a difference. Student `B` was *lucky*. The condition `B` associated with the patient's visible symptoms happened to be the condition the patient had, but if the professor had chosen a patient with c', student `B`

* In *New Directions in Logic, Language, and Computation*, eds. D. Lassiter and M. Slavkovic, Springer, 2012. The original publication is at www.springerlink.com.

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 that the patient's condition was c and passed the test.

Of course, A did not secure against *every* possibility of error. Suppose there is an extremely rare disease[1] 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 A did not know that the patient had c after all, since A did not rule out x?

According to a classic *relevant alternatives* style answer (e.g., [15, p. 775], [13, p. 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 a special reason to think that the patient may have had x instead of c, it should not have been necessary to rule out such a remote possibility in order to know that the patient had the common condition c.[2]

Much could be said about Example 1, but our interest here is in the pressure it appears to put on the claim that knowledge is *closed under known implication*. At its simplest, this is the claim that if an agent knows φ and knows that φ implies ψ, then she knows ψ: $(K\varphi \wedge K(\varphi \rightarrow \psi)) \rightarrow K\psi$, familiar as the K axiom of standard epistemic logic [19,14]. One obvious objection to K 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" [27,16] is to develop a good theoretical model of the knowledge of such agents. However, according to a different objection to K made famous in epistemology by Dretske [12] and Nozick [26] (and applicable to more sophisticated closure claims), knowledge would not be closed under known implication even for "ideally astute logicians" [12, p. 1010], who always put two and two together and come to believe all the consequences of what they know. It is this objection, not the logical omniscience problem, that is our starting point.

If one accepts the analysis at the end of Example 1, then one is close to denying K. For suppose A knows that if her patient has c, then he does not have x (because x confers immunity to c), (i) $K(c \rightarrow \neg x)$. Since 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 A knows that the patient has condition c, (iii) Kc, together (i) through (iii) violate the following instance of K: (iv) $(Kc \wedge K(c \rightarrow \neg x)) \rightarrow K\neg x$. To retain K, one must say either that A does not know that the patient has condition c after all (having not excluded x), or else that 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). While the second option threatens to commit us to problematic "easy

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

[2] Skeptics about medical knowledge may substitute one of the standard cases in the epistemology literature with a similar structure (see, e.g., [12, p. 1015], [13, p. 369]).

knowledge" [8], the first option threatens to commit us to radical skepticism about knowledge, given the inevitability of uneliminated possibilities of error.

Response 1 Dretske [12] and others [26,17] respond to the inconsistency of (i) through (iv), a version of the now standard "skeptical paradox" [7,9], by arguing that K is invalid, for reasons other than bounded rationality. Dretske's explanation of why K is invalid even for ideally astute logicians is in terms of his Relevant Alternatives (RA) Theory of knowledge [13]. According to 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 $c \rightarrow \neg x$, student `A` rules out certain relevant alternatives. In order to know $\neg x$, `A` must rule out certain relevant alternatives. However, 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 (or $K(c \rightarrow \neg x)$) to hold. But x *is* an alternative that must be ruled out in order for $K\neg x$ to hold. It is because the relevant alternatives may be different for what is in the antecedent and consequent of K that K is not valid in general.

Response 2 Against Response 1, Lewis [25] and others [7,9] attempt to explain away apparent closure failures by appeal to *epistemic contextualism*, the thesis that the truth values of knowledge attributions are context sensitive. According to Lewis's contextualist RA theory, in the context $\mathcal{C}$ of our conversation before we raised the possibility of the rare disease x, that possibility was irrelevant; so although `A` had not eliminated the possibility of x, we could truly say in $\mathcal{C}$ that `A` knew (at time t) that the patient's condition was c (Kc). However, by raising the possibility of x in our conversation, we changed the context to a new $\mathcal{C}'$ in which the uneliminated possibility of x was relevant. Hence we could truly say in $\mathcal{C}'$ that `A` did *not* know that the patient did not have x ($\neg K\neg x$), although `A` knew that x confers immunity to c ($K(c \rightarrow \neg x)$), which did not require ruling out x. Is this a violation of K in context $\mathcal{C}'$? It is not, because in $\mathcal{C}'$, unlike $\mathcal{C}$, we could *no longer* truly say that `A` knew (at t) that the patient's condition was c (Kc), given that `A` had not eliminated the newly relevant possibility of x. Moreover, Lewis argues that there is no violation of K in context $\mathcal{C}$ either:

> Knowledge *is* closed under implication.... Implication preserves truth—that is, it preserves truth in any given, fixed context. But if we switch contexts, all bets are off.... Dretske gets the phenomenon right...it is just that he misclassifies what he sees. He thinks it is a phenomenon of logic, when really it is a phenomenon of pragmatics. Closure, rightly understood, survives the rest. If we evaluate the conclusion for truth not with respect to the context in which it was uttered, but instead with respect to the different context in which the premise was uttered, then truth is preserved. (564)

Lewis claims that if we evaluate the consequent of (iv), $K\neg x$, with respect to the context $\mathcal{C}$ of our conversation before we raised the possibility of x, then it should come out *true*—despite the fact that `A` had not eliminated the possibility of x through any special tests—because the possibility of x was irrelevant in $\mathcal{C}$. If this is correct, then there is no violation of K in either context $\mathcal{C}'$ or $\mathcal{C}$.

This article introduces a formal framework to study Responses 1 and 2: in §2, the response of denying K leads to a study of other closure principles; in §3, the response of maintaining K with contextualism leads to a study of context dynamics. Having focused on the first response in detail elsewhere [20], here I focus more on the second, especially on logical issues associated with developing a *dynamic epistemic logic* [11,2] of context change over models for the RA theory.

2 Relevant Alternatives

An important distinction between versions of the RA theory, which our formalization will capture, has to do with logical structure. On the one hand, Dretske [13] 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 [17] considers (and rejects) an interpretation of the RA theory in which "there is a certain set of worlds selected as relevant," independently of any proposition, "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 worlds in which P is false, let $r(P)$ be the relevancy set for P, for which Dretske assumes $r(P) \subseteq \overline{P}$. To be more precise, since objective features of an agent's situation in world w may affect what alternatives are relevant (see [13, p. 377] and [10, p. 30f] on "subject factors"), let us write $r(P, w)$ for the relevancy set for P in world w, which may differ from $r(P, v)$ 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 (see [10, p. 30f] on "attributor factors"), then we should write $r_{\mathcal{C}}(P, w)$ to make the relativization to context explicit.

The quote from Dretske suggests the following definition:

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

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

> $\mathrm{RS}_{\exists\forall}$: for every context $\mathcal{C}$ and world w, there is ($\exists$) a set of *relevant* (*in* w) *worlds*, $R_{\mathcal{C}}(w)$, such that for every ($\forall$) proposition P, in order to know P in w (relative to $\mathcal{C}$) one must rule out the worlds in $R_{\mathcal{C}}(w) \cap \overline{P}$.

As a simple logical observation, every $\mathrm{RS}_{\exists\forall}$ theory is a $\mathrm{RS}_{\forall\exists}$ theory (take $r_{\mathcal{C}}(P, w) = R_{\mathcal{C}}(w) \cap \overline{P}$), but not necessarily *vice versa*. From now on, when I refer to $\mathrm{RS}_{\forall\exists}$ theories, I have in mind theories that are not also $\mathrm{RS}_{\exists\forall}$ theories. This distinction is at the heart of the disagreement about epistemic closure between Dretske and Lewis [25], as Lewis clearly adopts an $\mathrm{RS}_{\exists\forall}$ theory.

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" [18, p. 25] with "those that are too far away from the actual world being irrelevant" [17, p. 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 $\mathcal{M}'$ (see §3). 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 [10, p. 33]).

For simplicity (and in line with [25]) we will not represent in our RA models an agent's beliefs separately from her knowledge. Adding the usual machinery to do so is easy, but if the only point is to add *believing* φ as a necessary condition for knowing φ, it will not change any of our results about RA knowledge.

Definition 1 (RA Model). A *relevant alternatives model* is a tuple $\mathcal{M}$ of the form $\langle W, \rightarrow\!\!\!\!\triangleright, \preceq, V\rangle$ where:

1. W is a non-empty set;
2. $\rightarrow\!\!\!\!\triangleright$ 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) 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.

In addition, I assume the *well-foundedness* of each $\preceq_w$ (always satisfied in finite models) in what follows, since it allows us to state more perspicuous truth definitions. However, this does not affect our results about closure (see [20]).

I refer to elements of W as "worlds" or "possibilities" interchangeably. As usual, the function V maps each atom p to the set of worlds $V(p)$ where it holds.

Take $w \rightarrow\!\!\!\!\triangleright\ v$ to mean that v is an *uneliminated* possibility for the agent in w. According to Lewis's [25] notion of elimination, $\rightarrow\!\!\!\!\triangleright$ should be an equivalence relation; but for generality I assume only that $\rightarrow\!\!\!\!\triangleright$ is reflexive, reflecting the fact that an agent can never eliminate her actual world as a possibility. Whether we assume transitivity and symmetry in addition to reflexivity does not affect our results about closure, unless we make further assumptions about $\preceq$ (see [20]).

Take $u \preceq_w v$ to mean that u is *at least as relevant* (at w) as v is.[3] A relation satisfying Definition 1.3a is a *preorder*. The family of preorders in an RA model is like one of Lewis's (weakly centered) comparative similarity systems [23, §2.3] or standard γ-models [22], but without his assumption that each $\preceq_w$ is *total* on its field W_w. Condition 3b, that w is at least as relevant at w as any other

[3] 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 [23, §2.3] convention is now standard.

world is, follows Lewis's [25] *Rule of Actuality* that "actuality is always a relevant alternative" (554). Allowing $\preceq_w \neq \preceq_v$ when $w \neq v$ reflects the *world-relativity* of comparative relevance (based on "subject factors") mentioned 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. Moreover, we allow $\preceq_w \neq \preceq_v$ even when v is an uneliminated possibility for the agent in w, so $w \nrightarrow v$. For we do not assume that in w the agent can eliminate any v for which $\preceq_v \neq \preceq_w$. As Lewis [25] put it, "the subject himself may not be able to tell what is properly ignored" (554).

Notation 1 (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 $v \preceq_w u$;
- $\mathrm{Min}_{\preceq_w}(S) = \{v \in S \cap W_w \mid \text{there is no } u \in S \text{ such that } u \prec_w v\}$.

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. $\mathrm{Min}_{\preceq_w}(S)$ is the set of *most relevant* (at w) possibilities out of those in S that are ordered by $\preceq_w$.

When it comes to choosing a formal language to go with our RA models, we have a number of choices. For our first, we choose the following (cf. §3.2).

Definition 2 (Epistemic Language). Let $\mathsf{At} = \{p, q, r, \dots\}$ be a set of atomic sentences. The *epistemic language* is generated as follows, where $p \in \mathsf{At}$:

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

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 now interpret the language of Definition 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 supposed 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—in the spirit of Descartes' worries about error in the First Meditation. D-semantics is one (but not the only) way of understanding Dretske's [13] $\mathrm{RS}_{\forall\exists}$ theory, using Heller's [18,17] picture of relevance orderings of worlds.[4] Finally, L-semantics follows Lewis's [25] $\mathrm{RS}_{\exists\forall}$ theory (for a fixed context).

Definition 3 (Truth in an RA Model). Given a well-founded RA model $\mathcal{M} = \langle W, \nrightarrow, \preceq, V\rangle$ 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:

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

[4] Elsewhere [21] I argue for a better way of developing Dretske's [13] $\mathrm{RS}_{\forall\exists}$ theory, without the familiar world-ordering picture. Hence I take the 'D' in D-semantics as loosely as the 'C' in C-semantics. Still, 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 the K operator, the C-semantics clause is that of standard modal logic:

$$\mathcal{M}, w \vDash_c K\varphi \text{ iff } \forall v \in W\text{: if } w \rightarrow\!\!\!\!\triangleright\ v \text{ then } \mathcal{M}, v \vDash_c \varphi,$$

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 2 (Extension and Complement) Where $\mathcal{M} = \langle W, \rightarrow\!\!\!\!\triangleright, \preceq, V\rangle$,

- $[\![\varphi]\!]^{\mathcal{M}}_x = \{v \in W \mid \mathcal{M}, v \vDash_x \varphi\}$ 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, we write $[\![\varphi]\!]$.
- For $S \subseteq W$, we write $\overline{S} = \{v \in W \mid v \notin S\}$ for the complement of S in W.

Definition 4 (Truth in an RA Model cont.). For C-, D-, and L-semantics, the clauses for the K operator are:

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

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

If φ is valid in X-semantics, we say that φ is *X-valid* and write $\vDash_x \varphi$.

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.

It is easy to check that according to C/D/L-semantics, whatever is known is true. For D- and L-semantics, Fact 1 reflects Lewis's [25, p. 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 $\preceq_w$ (Definition 1.3b) and $w \rightarrow\!\!\!\!\triangleright\ w$.

Fact 1 (Veridicality) $K\varphi \rightarrow \varphi$ is C/D/L-valid.

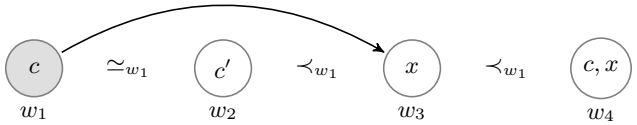

Fig. 1: an RA model for Example 1 (partially drawn, reflexive loops omitted)

Consider the model in Fig. 1, drawn for student A in Example 1. An arrow from w to v indicates that $w \twoheadrightarrow v$. (For all $v \in W$, $v \twoheadrightarrow v$, but we omit all reflexive loops.) The ordering of the worlds by their relevance at w_1, thought of as the actual world, is indicated between worlds.[5] In w_1, the patient has the common condition c, represented by the atomic sentence c true at w_1. 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.[6] 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 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 $K(\neg x \rightarrow c)$ is true at w_1 in C-semantics.

Consider D-semantics. First observe that D-semantics issues our original verdict that student A knows 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 A in w_1. $K(c \rightarrow \neg x)$ is also true at w_1, since the most relevant (at w_1) $\neg(c \rightarrow \neg x)$-world, w_4, is ruled out by 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 A in w_1. However, the most relevant (at w_1) x-world, w_3, is *not* ruled out by A in w_1, so $K\neg x$ is false at w_1 in D-semantics. Hence A does not know that the patient does not have x.

We have shown the second part of the following fact, which matches Dretske's [12] view. The first part, which is standard, matches Lewis's [25, p. 563n21].

Fact 2 (Known Implication) The principles $K\varphi \wedge K(\varphi \rightarrow \psi) \rightarrow K\psi$ and $K\varphi \wedge K(\varphi \leftrightarrow \psi) \rightarrow K\psi$ are C/L-valid, but not D-valid.

Finally, consider the model in Fig. 1 from the perspective of L-semantics. What is noteworthy in this case is that according to L-semantics, student A *does* know that the patient does not have disease x. $K\neg x$ is true at w_1, because $\neg x$ is true in all of the most relevant (at w_1) worlds, namely in w_1 and w_2.

In the terminology of Dretske [12, p. 1007], Fact 2 shows that the knowledge operator K is not *fully penetrating*, since it does not penetrate to all logical consequence of what is known. Yet Dretske claims that K is *semi-penetrating*, since

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

[6] We could add new atomic sentences t_c and $t_{c'}$ standing for "the test results favor c over c'" and "the test results favor c' over c," respectively. We would then make t_c true and $t_{c'}$ false at w_1, w_3, and w_4, while making $t_{c'}$ true and t_c false at w_2.

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, even these monotonicity principles fail.

Fact 3 (Simplification & Addition) The principles $K(\varphi \wedge \psi) \to K\varphi \wedge K\psi$ and $K\varphi \to K(\varphi \vee \psi)$ are C/L-valid, but not D-valid.

Proof The proof of C/L-validity is standard. For D-semantics, the pointed model $\mathcal{M}, w_1$ in Fig. 1 falsifies both $K(c \wedge \neg x) \to K\neg x$ 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. □

Facts 2 and 3 point 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. 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. It is difficult to escape this dilemma while retaining something like Heller's [18,17] world-ordering picture with which we started before Definition 1. In [21], I propose a different way of developing the theory such that the knowledge operator is semi-penetrating in Dretske's sense, thereby avoiding the dilemma above.

Facts 2 and 3 also raise the question: what is the complete logic of knowledge over RA models? Theorem 1, proven in [20,21], gives the answer. Interestingly, the answer depends on whether we assume that each $\preceq_w$ is *total* on its field W_w ($\forall u, v \in W_w$: $u \preceq_w v$ or $v \preceq_w u$), so that $\preceq_w$ is a *ranking* of worlds in W_w by their relevance. (In this case, we call the RA model itself "total.") Following the nomenclature of Chellas [6], **E** is the weakest of the classical modal systems extending propositional logic with the rule RE, and $\mathbf{ES}_1 \ldots \mathbf{S}_n$ is the extension of **E** with every instance of schemas $\mathrm{S}_1 \ldots \mathrm{S}_n$. The X axiom schema is new.

RE. $\dfrac{\varphi \leftrightarrow \psi}{K\varphi \leftrightarrow K\psi}$ T. $K\varphi \to \varphi$ N. $K\top$

C. $K\varphi \wedge K\psi \to K(\varphi \wedge \psi)$ M. $K(\varphi \wedge \psi) \to K\varphi \wedge K\psi$ X. $K(\varphi \wedge \psi) \to K\varphi \vee K\psi$

Theorem 1 (Completeness).

1. *The system* ***EMCNT*** *(****KT****) is sound and complete for C/L-semantics over RA models.*
2. (The Logic of Ranked Relevant Alternatives) *The system* ***ECNTX*** *is sound and complete for D-semantics over* total *RA models.*
3. *The system* ***ECNT*** *is sound and complete for D-semantics over RA models.*

3 The Dynamics of Context

In this section, we extend our formalization to capture the *contextualist* Response 2 to Example 1 in §1. (It may be helpful to reread Response 2 as a reminder.)

In the framework of Lewis [24], the family $\preceq$ of relevance orderings in an RA model may be thought of as a component of the *conversational score*. Changes in this component of the conversational score, an aspect of what Lewis calls the *kinematics of score*, correspond to transformations of RA models. We begin with an RA model $\mathcal{M}$ representing what an agent counts as knowing relative to an initial conversational context. If some change in the conversation makes the issue of φ relevant, then corresponding to this change the model transforms from $\mathcal{M}$ to $\mathcal{M}^{\uparrow\varphi}$. In the new model, what the agent counts as knowing may be different.

For variety, we will define two types of operations on models, $\uparrow\varphi$ and $\nmid\varphi$. Roughly speaking, $\uparrow\varphi$ changes the model so that the *most relevant φ-worlds* in $\mathcal{M}$ become among the *most relevant worlds overall* in $\mathcal{M}^{\uparrow\varphi}$. By contrast, $\nmid\varphi$ changes the model so that any worlds *at least as relevant as* the most relevant φ-worlds in $\mathcal{M}$ become among the most relevant worlds overall in $\mathcal{M}^{\nmid\varphi}$. The following definition makes these descriptions more precise. For convenience, in this section we assume that each preorder $\preceq_w$ is total on its field W_w, but all of the definitions and results can be modified to apply to the non-total case.

Definition 5 (RA Context Change). Given an RA model $\mathcal{M} = \langle W, \rightarrow, \preceq, V\rangle$, define the models $\mathcal{M}^{\uparrow\varphi} = \langle W, \rightarrow, \preceq^{\uparrow\varphi}, V\rangle$ and $\mathcal{M}^{\nmid\varphi} = \langle W, \rightarrow, \preceq^{\nmid\varphi}, V\rangle$ such that for all $w, u, v \in W$:

1. if $u \in \mathrm{Min}_{\preceq_w}\left([\![\varphi]\!]^{\mathcal{M}}\right) \cup \mathrm{Min}_{\preceq_w}(W)$, then $u \preceq^{\uparrow\varphi}_w v$;
2. if $u, v \notin \mathrm{Min}_{\preceq_w}\left([\![\varphi]\!]^{\mathcal{M}}\right) \cup \mathrm{Min}_{\preceq_w}(W)$, then $u \preceq^{\uparrow\varphi}_w v$ iff $u \preceq_w v$;

and

3. if $\exists x \in \mathrm{Min}_{\preceq_w}\left([\![\varphi]\!]^{\mathcal{M}}\right)$ such that $u \preceq_w x$, then $u \preceq^{\nmid\varphi}_w v$;
4. if $\forall x \in \mathrm{Min}_{\preceq_w}\left([\![\varphi]\!]^{\mathcal{M}}\right)$, $u \not\preceq_w x$ and $v \not\preceq_w x$, then $u \preceq^{\nmid\Phi}_w v$ iff $u \preceq_w v$.

In other words, for $\uparrow\varphi$, the *most relevant φ-worlds* according to $\preceq_w$ become among the *most relevant worlds* according to $\preceq^{\uparrow\varphi}_w$; the most relevant worlds according to $\preceq_w$ remain among the most relevant worlds according to $\preceq^{\uparrow\varphi}_w$; and for all other worlds, $\preceq^{\uparrow\varphi}_w$ agrees with $\preceq_w$. For $\nmid\varphi$, all worlds *at least as relevant as* the most relevant φ-worlds according to $\preceq_w$ become among the most relevant worlds according to $\preceq^{\nmid\varphi}_w$; and for all other worlds, $\preceq^{\nmid\varphi}_w$ agrees with $\preceq_w$.

Which of these operations is most appropriate for modeling a given context change is an interesting question, which I leave aside here. Other operations could be defined as well, but these will suffice as examples of the general method. Fig. 2 shows the application of either $\uparrow x$ or $\nmid x$ (denoted $+x$) to the model $\mathcal{M}$ for Example 1, the result of which is the same for both. Fig. 3 shows $\uparrow x$ and $\nmid x$ applied to a different initial model, $\mathcal{N}$, in which case the results are different.

To describe the effect of these context change operations using our formal language, we extend the language of Definition 2 with *dynamic* context change

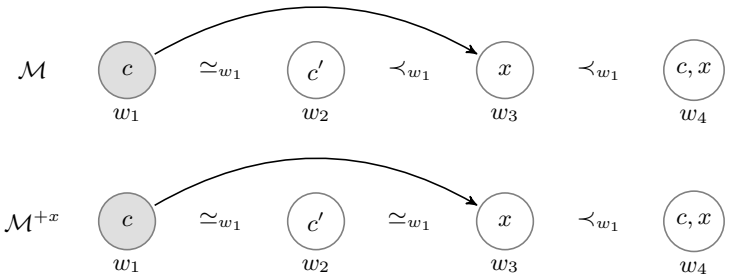

Fig. 2: result of context change by raising the possibility of x in Example 1

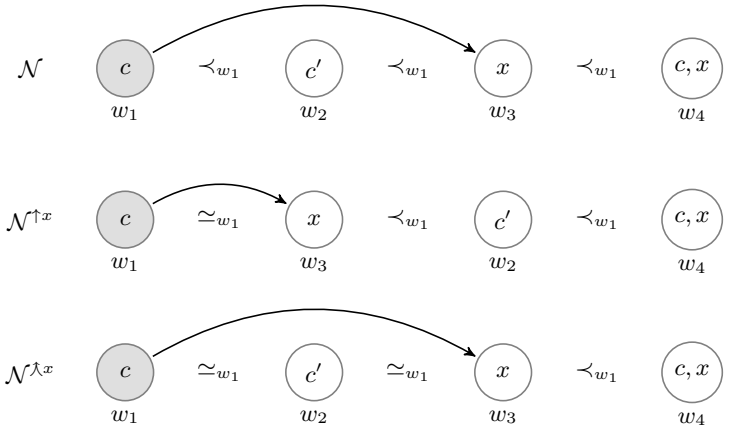

Fig. 3: different results of context change by $\uparrow x$ and $\nuparrow x$

operators of the form $[+\varphi]$ for $+ \in \{\uparrow, \nuparrow\}$, in the style of dynamic epistemic logic [11,2]. One can read $[+\varphi]\psi$ as "after φ becomes relevant, ψ is the case" or "after φ is raised, ψ is the case" or "after context change by φ, ψ is the case," etc.

Definition 6 (Contextualist Epistemic Language). Let $\mathsf{At} = \{p, q, r, \dots\}$ be a set of atomic sentences. The *contextualist epistemic language* is generated as follows, where $p \in \mathsf{At}$:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K\varphi \mid [\pi]\varphi$$
$$\pi ::= \uparrow \varphi \mid \nuparrow \varphi.$$

We give the truth clauses for the operators $[\uparrow\ \varphi]$ and $[\nuparrow\ \varphi]$ with the help of Definition 5, using $+$ to stand for either $\uparrow$ or $\nuparrow$ in definitions applicable to both.

Definition 7 (Truth). The truth clause for the context change operator is:

$$\mathcal{M}, w \vDash [+\varphi]\psi \text{ iff } \mathcal{M}^{+\varphi}, w \vDash \psi.$$

In other words, "after context change by φ, ψ is the case" is true at w in the initial model $\mathcal{M}$ if and only if ψ is true at w in the new model $\mathcal{M}^{+\varphi}$.

Having set up this contextualist machinery, there are a number of directions to explore. Given the space available here, we will touch on two: first, a brief comparison between (non-contextualist) D-semantics and contextualist L-semantics; second, a technical excursion in search of *reduction axioms* for context change.

3.1 D-Semantics vs. Contextualist L-Semantics

The following fact matches Lewis's [25] view on closure and context from §1.

Fact 4 (Known Implication Cont.) According to D-semantics, closure under known implication can fail. According to L-semantics, closure under known implication always holds for a fixed context, but may fail across context changes:

1. $\not\models_d K\varphi \wedge K(\varphi \to \psi) \to K\psi$
2. $\models_l K\varphi \wedge K(\varphi \to \psi) \to K\psi$
3. $\not\models_l K\varphi \wedge K(\varphi \to \psi) \to [+\neg\psi]K\psi$
4. $\not\models_l K\varphi \to [+\neg\psi](K(\varphi \to \psi) \to K\psi)$

Proof We have already noted part 1 and 2 in §2. For 3, its instance

$$Kc \wedge K(c \to \neg x) \to [+x]K\neg x \tag{1}$$

is false at $\mathcal{M}, w_1$ in Fig. 2. As we saw in §2, the antecedent is true at $\mathcal{M}, w_1$. To determine whether $\mathcal{M}, w_1 \models_l [+x]K\neg x$, by Definition 7 we must check whether $\mathcal{M}^{+x}, w_1 \models_l K\neg x$. Since in $\mathcal{M}^{+x}$ there is a most relevant (at w_1) world, w_3, which satisfies x and is not ruled out at w_1, we have $\mathcal{M}^{+x}, w_1 \not\models_l K\neg x$. Therefore, $\mathcal{M}, w_1 \not\models_l [+x]K\neg x$, so (1) is false at $\mathcal{M}, w_1$. It is also easy to check that $\mathcal{M}^{+x}, w_1 \models K(c \to \neg x)$, so the corresponding instance of 4 is false at $\mathcal{M}, w_1$. □

We will use the next fact to generalize Fact 4 to all kinds of closure failure (Fact 6), not only failures of closure under known implication.

Fact 5 (Relation of D- to Contextualist L-semantics) Given an RA model $\mathcal{M} = \langle W, \rightarrow, \preceq, V\rangle$ with $w \in W$, for any *propositional* formula φ,

$$\mathcal{M}, w \models_d K\varphi \text{ iff } \mathcal{M}, w \models_l [+\neg\varphi]K\varphi.$$

Proof For the case where $+$ is $\uparrow$, by Definition 5,

$$\mathrm{Min}_{\preceq_w^{\uparrow\neg\varphi}}(W) = \mathrm{Min}_{\preceq_w}(W) \cup \mathrm{Min}_{\preceq_w}(\overline{[\![\varphi]\!]}^{\mathcal{M}}), \tag{2}$$

so

$$\mathrm{Min}_{\preceq_w^{\uparrow\neg\varphi}}(W) \cap \overline{[\![\varphi]\!]}^{\mathcal{M}^{\uparrow\neg\varphi}} = (\mathrm{Min}_{\preceq_w}(W) \cup \mathrm{Min}_{\preceq_w}(\overline{[\![\varphi]\!]}^{\mathcal{M}})) \cap \overline{[\![\varphi]\!]}^{\mathcal{M}^{\uparrow\neg\varphi}}. \tag{3}$$

Since φ is propositional, by an obvious induction we have

$$\overline{[\![\varphi]\!]}^{\mathcal{M}^{+\neg\varphi}} = \overline{[\![\varphi]\!]}^{\mathcal{M}}, \tag{4}$$

so from (3) we have

$$\begin{aligned}\mathrm{Min}_{\preceq_w^{\uparrow\neg\varphi}}(W) \cap \overline{[\![\varphi]\!]}^{\mathcal{M}^{\uparrow\neg\varphi}} &= (\mathrm{Min}_{\preceq_w}(W) \cup \mathrm{Min}_{\preceq_w}(\overline{[\![\varphi]\!]}^{\mathcal{M}})) \cap \overline{[\![\varphi]\!]}^{\mathcal{M}} \\ &= \mathrm{Min}_{\preceq_w}(\overline{[\![\varphi]\!]}^{\mathcal{M}}). \end{aligned} \tag{5}$$

It follows from (5) that

$$\forall v \in \mathrm{Min}_{\preceq_w}(\overline{[\![\varphi]\!]}^{\mathcal{M}})\colon w \not\rightarrow v \tag{6}$$

is equivalent to

$$\forall v \in \mathrm{Min}_{\preceq_w^{\uparrow\neg\varphi}}(W) \cap \overline{[\![\varphi]\!]}^{\mathcal{M}^{\uparrow\neg\varphi}} : w \not\rightarrow v, \tag{7}$$

which by Definition 4 means that $\mathcal{M}, w \models_d K\varphi$ is equivalent to $\mathcal{M}^{\uparrow\neg\varphi}, w \models_l K\varphi$, which by Definition 7 is equivalent to $\mathcal{M}, w \models_l [\uparrow \neg\varphi]K\varphi$.

The proof for the case where $+$ is $\nwarrow$ is similar. □

Using Fact 5, we can now state a generalization of Fact 4 as follows.

Fact 6 (Inter-context Closure Failure) Let $\varphi_1, \ldots, \varphi_n$ and ψ be propositional formulas. Given an RA model $\mathcal{M} = \langle W, \rightarrow, \preceq, V\rangle$ with $w \in W$, if

$$\mathcal{M}, w \not\models_d K\varphi_1 \wedge \cdots \wedge K\varphi_n \rightarrow K\psi$$

then

$$\mathcal{M}, w \not\models_l K\varphi_1 \wedge \cdots \wedge K\varphi_n \rightarrow [+\neg\psi]K\psi.$$

Proof Assume the first line. Since for any formula φ, $\mathcal{M}, w \models_d K\varphi$ implies $\mathcal{M}, w \models_l K\varphi$, we have $\mathcal{M}, w \models_l K\varphi_1 \wedge \cdots \wedge K\varphi_n$. Since $\mathcal{M}, w \not\models_d K\psi$, we have $\mathcal{M}, w \not\models_l [+\neg\psi]K\psi$ by Fact 5, which gives the second line. □

Most contextualists deny that closure fails in any of the ways allowed by D-semantics. But Fact 6 shows that for *every way* in which closure fails for D-semantics, there is a corresponding *inter-context* "closure failure" for L-semantics when the context changes with the negation of the consequent of the closure principle becoming relevant. According to some standard contextualist views, asserting that the agent knows the consequent has just this effect on the context. For example, according to DeRose [9], "When it's asserted that S knows (or doesn't know) that P, then, if necessary, enlarge the sphere of epistemically relevant worlds so that it at least includes the closest worlds in which P is false" (37). According to Lewis [25], "No matter how far-fetched a certain possibility may be, no matter how properly we might have ignored it in some other context, if in this context we are not in fact ignoring it but attending to it, then for us now it is a relevant alternative" (559). If such views of the shiftiness of context are correct, then Fact 6 shows that contextualists who claim to "preserve closure"—with respect to a fixed context—may not vindicate *closure reasoning* (reasoning over time about an agent's knowledge that applies closure principles to draw conclusions) any more than those who allow failures of closure as in D-semantics.

Much more could be said about these conceptual issues (see [21]), but now we will pursue a different line, checking our logical grip on the dynamics of context.

3.2 Reduction Axioms for Context Change

In this section, we turn to a more technical topic. Our goal is to apply one of the main ideas of dynamic epistemic logic, that of *reduction axioms*, to the picture of context change presented in the previous sections. Roughly speaking, reduction axioms are valid equivalences of the form $[+\chi]\psi \leftrightarrow \psi'$, where the left-hand side states that some ψ is true *after* the context change with χ, while the right-hand side gives an *equivalent* ψ' describing what is true *before* the context change. For example, we can ask whether the agent counts as knowing φ after χ becomes relevant, i.e., is $[+\chi]K\varphi$ true? The reduction axioms will answer this question by describing what must be true of the agent's epistemic state *before* the context change in order for the agent to count as knowing φ after the context change.

To obtain reduction axioms for context change that are valid over our RA models, we will use a language more expressive than the epistemic language used in the previous sections. Our new *RA language* will be capable of describing what is relevant at a world and what is ruled out at a world independently. This additional expressive power will allow us to obtain reduction axioms using methods similar to those applied by van Benthem and Liu [4] to *dynamic epistemic preference logic* (also see [3]), but with an important difference.

Van Benthem and Liu work with models with a *single* preorder over worlds (for each agent), representing an agent's preferences between worlds, and their language contains an operator $\Box^{\succ}$ used to quantify over all worlds that are *better* than the current world according to the agent.[7] In our setting, $\Box^{\succ}$ would quantify over all worlds that are *more relevant*. Using another operator $\Box^{\rightarrow}$ to quantify over all worlds that are *uneliminated* at the current world, we can try to write a formula expressing that all of the most relevant $\neg\varphi$-worlds are eliminated at the current world. An equivalent statement is that for all uneliminated worlds v, if v is a $\neg\varphi$-world, then there is another $\neg\varphi$-world that is strictly more relevant than v. This is expressed by $\Box^{\rightarrow}(\neg\varphi \rightarrow \Diamond^{\succ}\neg\varphi)$, where $\Diamond^{\succ}\psi := \neg\Box^{\succ}\neg\psi$.

The problem with the above approach is that unlike the models of van Benthem and Liu (but like models for conditional logic and the general *belief revision structures* of [5]), our RA models include a preorder $\preceq_w$ for *each* world w. Hence if the operator $\Box^{\succ}$ quantifies over all worlds that are more relevant than the current world according to the relevance relation of the current world, then $\Box^{\rightarrow}(\neg\varphi \rightarrow \Diamond^{\succ}\neg\varphi)$ will be true at w just in case for all worlds v uneliminated at w, if v is a $\neg\varphi$-world, then there is another $\neg\varphi$-world that is strictly more relevant than v *according to* $\preceq_v$. Yet this is not the desired truth condition.[8] The desired truth condition is that for all worlds v uneliminated at w, if v is a $\neg\varphi$-world, then there is another $\neg\varphi$-world that is strictly more relevant than

[7] Van Benthem et al. [3] write this operator as $\Box^{<}$, since they take $w \prec v$ to mean that v is strictly better than w according to the agent. Since we take $w \prec v$ to mean that w is strictly more relevant than v, we write $\Box^{\succ}$ for the operator that quantifies over more relevant worlds. We will write $\Box^{\preceq}$ for the operator that quantifies over worlds that are of equal or lesser relevance. We use the same $\preceq$ for the superscript of the operator and for the relation in the model, trusting that no confusion will arise.

[8] Since v is assumed to be minimal in $\preceq_v$, the condition would never be met.

v *according to* $\preceq_w$. To capture this truth condition, we will use an approach inspired by *hybrid logic* [1]. First, different modalities $\Box^{\succ_x}$, $\Box^{\succ_y}$, etc., will be associated in a given model with different relevance relations $\preceq_w$, $\preceq_v$, etc., by an assignment function g. Second, a *binder* $\downarrow$ will be used to bind a world variable x to the current world, so that the formula $\downarrow x.\Box^{\rightarrow}(\neg\varphi \rightarrow \Diamond^{\succ_x}\neg\varphi)$ will capture the desired truth condition described above (cf. [23, §2.8] on the † operator).

In addition to the operator $\Box^{\succ_x}$ that quantifies over all worlds more relevant than the current world according to $\preceq_{g(x)}$, we will use an operator $\Box^{\preceq_x}$ that quantifiers over all worlds whose relevance is equal to or lesser than that of the current world according to $\preceq_{g(x)}$. The second operator is necessary for writing reduction axioms for the context change operation $\nwarrow\!\!\!\!\!\!\uparrow$. Together the two types of operators will also allow us to quantify over all worlds in the field of $\preceq_{g(x)}$, $W_{g(x)}$, with formulas of the form $\Box^{\succ_x}\varphi \wedge \Box^{\preceq_x}\varphi$, which we will use in writing reduction axioms for both of the context change operations, $\uparrow$ and $\nwarrow\!\!\!\!\!\!\uparrow$.

Definition 8 (Dynamic & Static RA Languages). Let $\mathsf{At} = \{p, q, r \dots\}$ be a set of atomic sentences and $\mathsf{Var} = \{x, y, z, \dots\}$ a set of variables. The *dynamic RA language* is generated as follows, where $p \in \mathsf{At}$ and $x \in \mathsf{Var}$:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box^{\rightarrow}\varphi \mid \Box^{\preceq_x}\varphi \mid \Box^{\succ_x}\varphi \mid \downarrow x.\varphi \mid [\pi]\varphi$$
$$\pi ::= \uparrow \varphi \mid \nwarrow\!\!\!\!\!\!\uparrow \varphi.$$

Where R is $\preceq_x$, $\succ_x$, or $\rightarrow$, let $\Diamond^R\varphi := \neg\Box^R\neg\varphi$; let R_x stand for either $\preceq_x$ or $\succ_x$ in definitions that apply to both; and let us use $+$ as before. Finally, let the *static RA language* be the fragment of the dynamic RA language consisting of those formulas that do not contain any context change operators $[\pi]$.

The truth clauses are as one would expect from our description above, and the clause for the context change operators is the same as Definition 7.

Definition 9 (Truth). Given an RA model $\mathcal{M} = \langle W, \rightarrow, \preceq, V\rangle$ and an assignment function $g\colon \mathsf{Var} \to W$, we define $\mathcal{M}, g, w \vDash \varphi$ as follows (with propositional cases as in Definition 3):

$$\begin{array}{lll}
\mathcal{M}, g, w \vDash \Box^{\rightarrow}\varphi & \text{iff} & \forall v \in W\text{: if } w \rightarrow v \text{ then } \mathcal{M}, g, v \vDash \varphi;\\
\mathcal{M}, g, w \vDash \Box^{R_x}\varphi & \text{iff} & \forall v \in W\text{: if } wR_{g(x)}v \text{ then } \mathcal{M}, g, v \vDash \varphi;\\
\mathcal{M}, g, w \vDash [+\chi]\varphi & \text{iff} & \mathcal{M}^{+\chi}, g, w \vDash \varphi;\\
\mathcal{M}, g, w \vDash \downarrow x.\varphi & \text{iff} & \mathcal{M}, g^x_w, w \vDash \varphi,
\end{array}$$

where g^x_w is such that $g^x_w(x) = w$ and $g^x_w(y) = g(y)$ for all $y \neq x$.

Hence the $\downarrow x.\varphi$ clause captures the idea of letting x stand for the current world by changing the assignment g to one that maps x to w but is otherwise the same.

We now show how the epistemic language can be translated into the RA language in two different ways, corresponding to D- and L-semantics.[9] To simplify the translation, let us assume for the moment that all of our RA models $\mathcal{M} = \langle W, \rightarrow, \preceq, V\rangle$ are *universal* in the sense that for all $w \in W$, $W_w = W$.

[9] Note that since the translation of Definition 10 only requires a single variable x, for our purposes here it would suffice to define the RA language such that $|\mathsf{Var}| = 1$.

Definition 10 (Translation). Let σ_d be a translation from the epistemic language of Definition 2 to the static RA language of Definition 8 defined by:

$$\begin{aligned}\sigma_d(p) &= p\\ \sigma_d(\neg\varphi) &= \neg\sigma_d(\varphi)\\ \sigma_d(\varphi\wedge\psi) &= (\sigma_d(\varphi)\wedge\sigma_d(\psi))\\ \sigma_d(K\varphi) &= \downarrow x.\Box^{\rightarrow\!\!\!\triangleright}(\neg\sigma_d(\varphi)\rightarrow\Diamond^{\succ_x}\neg\sigma_d(\varphi)).\end{aligned}$$

Let σ_l be a translation analogous to σ_d but with

$$\sigma_l(K\varphi) = \downarrow x.\Box^{\rightarrow\!\!\!\triangleright}(\neg\sigma_l(\varphi)\rightarrow\Diamond^{\succ_x}\top).$$

As explained at the beginning of this section, the idea of the σ_d translation is that the truth clause for $K\varphi$ in D-semantics—stating that the most relevant $\neg\varphi$-worlds are eliminated—is equivalent to the statement that for all worlds v uneliminated at the current world w, if v is a $\neg\varphi$-world, then there is another $\neg\varphi$-world that is strictly more relevant than v according to $\preceq_w$. This is exactly what $\sigma_d(K\varphi)$ expresses. Similarly, the idea of the σ_l translation is that the truth clause for $K\varphi$ in L-semantics—stating that among the most relevant worlds overall, all $\neg\varphi$-worlds are eliminated—is equivalent to the statement that for all worlds v uneliminated at the current world w, if v is a $\neg\varphi$-world, then there is another world that is strictly more relevant than v according to $\preceq_w$, in which case v is not among the most relevant worlds overall according to $\preceq_w$. This is exactly what $\sigma_l(K\varphi)$ expresses. The following proposition confirms these claims.

Proposition 1 (Simulation). For any RA model $\mathcal{M} = \langle W, \rightarrow\!\!\!\triangleright, \preceq, V\rangle$, assignment $g\colon \mathsf{Var}\to W$, world $w\in W$, and formula φ of the epistemic language:

$$\begin{aligned}&\mathcal{M}, w \vDash_d \varphi \;\text{ iff }\; \mathcal{M}, g, w \vDash \sigma_d(\varphi);\\ &\mathcal{M}, w \vDash_l \varphi \;\text{ iff }\; \mathcal{M}, g, w \vDash \sigma_l(\varphi).\end{aligned}$$

Proof By induction on φ. All of the cases are trivial except where φ is of the form $K\psi$. By Definition 10, we are to show

$$\mathcal{M}, w \vDash_d K\psi \text{ iff } \mathcal{M}, g, w \vDash \downarrow x.\Box^{\rightarrow\!\!\!\triangleright}(\neg\sigma_d(\varphi)\rightarrow\Diamond^{\succ_x}\neg\sigma_d(\varphi)). \tag{8}$$

By Definition 9, the rhs of (8) holds iff for all $v\in W$, if $w\rightarrow\!\!\!\triangleright\, v$, then

$$\mathcal{M}, g^x_w, v \vDash \neg\sigma_d(\psi)\rightarrow\Diamond^{\succ_x}\neg\sigma_d(\psi). \tag{9}$$

By Definition 9, (9) is equivalent to the disjunction of the following:

$$\mathcal{M}, g^x_w, v \vDash \sigma_d(\psi); \tag{10}$$

$$\exists u\in W\colon u \prec_{g^x_w(x)} v \text{ and } \mathcal{M}, g^x_w, u \nvDash \sigma_d(\psi). \tag{11}$$

By the inductive hypothesis, (10) and (11) are respectively equivalent to

$$\mathcal{M}, v \vDash_d \psi \text{ and} \tag{12}$$

$$\exists u\in W\colon u \prec_w v \text{ and } \mathcal{M}, u \nvDash_d \psi. \tag{13}$$

Assuming $\mathcal{M}$ is universal, the disjunction of (12) and (13) is equivalent to

$$v \notin \mathrm{Min}_{\preceq_w}(\overline{[\![\psi]\!]}). \tag{14}$$

Hence the rhs of (8) holds if and only if for all $v \in W$, if $w \twoheadrightarrow v$, then (14) holds. The rhs of this biconditional is equivalent to the lhs of (8), $\mathcal{M}, w \vDash_d K\psi$, by Definition 3. The proof for the case of L-semantics is similar. □

If we do not assume that RA models are universal, then we must modify the translation of Definition 10 such that

$$\sigma'_d(K\varphi) = \downarrow x.\Box^{\twoheadrightarrow}(\neg\sigma'_d(\varphi) \to (\Diamond^{\succ_x}\neg\sigma'_d(\varphi) \vee \Box^{\preceq_x}\bot));$$
$$\sigma'_l(K\varphi) = \downarrow x.\Box^{\twoheadrightarrow}(\neg\sigma'_l(\varphi) \to (\Diamond^{\succ_x}\top \vee \Box^{\preceq_x}\bot)).$$

We leave it to the reader to verify that given the modified translation, Proposition 1 holds for RA models that are not necessarily universal.

We are now prepared to do what we set out to do at the beginning of this section: give reduction axioms for the context change operations of Definition 5. For the following proposition, let us define $\Box^x\varphi := \Box^{\succ_x}\varphi \wedge \Box^{\preceq_x}\varphi$.

Proposition 2 (RA Reduction). Given the following valid reduction axioms and the rule of replacement of logical equivalents,[10] any formula of the *dynamic* RA language is equivalent to a formula of the *static* RA language:

$$[+\chi]\,p \leftrightarrow p; \tag{15}$$
$$[+\chi]\,\neg\varphi \leftrightarrow \neg\,[+\chi]\,\varphi; \tag{16}$$
$$[+\chi]\,(\varphi \wedge \psi) \leftrightarrow [+\chi]\,\varphi \wedge [+\chi]\,\psi; \tag{17}$$
$$[+\chi]\downarrow x.\varphi \leftrightarrow \downarrow x.[+\chi]\varphi; \tag{18}$$
$$[+\chi]\,\Box^{\twoheadrightarrow}\varphi \leftrightarrow \Box^{\twoheadrightarrow}\,[+\chi]\,\varphi; \tag{19}$$
$$\begin{aligned}[\uparrow\chi]\,\Box^{\succ_x}\varphi \leftrightarrow\ & \Box^{\succ_x}\bot \vee (\chi \wedge \Box^{\succ_x}\neg\chi)\\ & \vee \big(\Box^{\succ_x}[\uparrow\chi]\varphi \wedge \Box^{\preceq_x}((\chi \wedge \Box^{\succ_x}\neg\chi) \to [\uparrow\chi]\varphi)\big);\end{aligned} \tag{20}$$
$$\begin{aligned}[\uparrow\chi]\,\Box^{\preceq_x}\varphi \leftrightarrow\ & \big((\Box^{\succ_x}\bot \vee (\chi \wedge \Box^{\succ_x}\neg\chi)) \wedge \Box^x[\uparrow\chi]\varphi\big)\\ & \vee\ \Box^{\preceq_x}((\chi \wedge \Box^{\succ_x}\neg\chi) \vee [\uparrow\chi]\varphi);\end{aligned} \tag{21}$$
$$\begin{aligned}[\Uparrow\chi]\,\Box^{\succ_x}\varphi \leftrightarrow\ & \Diamond^{\preceq_x}(\chi \wedge \Box^{\succ_x}\neg\chi)\\ & \vee \big(\neg\Diamond^{\preceq_x}(\chi \wedge \Box^{\succ_x}\neg\chi) \wedge \Box^{\succ_x}\,[\Uparrow\chi]\,\varphi\big);\end{aligned} \tag{22}$$
$$\begin{aligned}[\Uparrow\chi]\,\Box^{\preceq_x}\varphi \leftrightarrow\ & \big(\Diamond^{\preceq_x}(\chi \wedge \Box^{\succ_x}\neg\chi) \wedge \Box^x[\Uparrow\chi]\varphi\big)\\ & \vee \big(\neg\Diamond^{\preceq_x}(\chi \wedge \Box^{\succ_x}\neg\chi) \wedge \Box^{\preceq_x}\,[\Uparrow\chi]\,\varphi\big).\end{aligned} \tag{23}$$

Proof Assuming the axioms are valid, the argument for the claim of the proposition is straightforward. Each of the axioms drives the context change

[10] Semantically, if $\alpha \leftrightarrow \beta$ is valid, so is $\varphi(\alpha/p) \leftrightarrow \varphi(\beta/p)$, where (ψ/p) indicates substitution of ψ for p.

operators $[+\chi]$ inward until eventually these operators apply only to atomic sentences p, at which point they can be eliminated altogether using (15). In case we encounter something of the form $[+\chi_1][+\chi_2]\varphi$, we first reduce $[+\chi_2]\varphi$ to an equivalent static formula φ' and then use the replacement of logical equivalents to obtain $[+\chi_1]\varphi'$, which we then reduce to an equivalent static formula φ'', etc.

Let us now check the validity of (15) - (19) in turn. First, (15) is valid because the context change operations of Definition 5 do not change the valuation V for atomic sentences in the model. For (16), in the left-to-right direction we have the following implications: $\mathcal{M}, w \vDash [+\chi]\neg\varphi \Rightarrow \mathcal{M}^{+\chi}, w \vDash \neg\varphi \Rightarrow \mathcal{M}^{+\chi}, w \nvDash \varphi \Rightarrow \mathcal{M}, w \nvDash [+\chi]\varphi \Rightarrow \mathcal{M}, w \vDash \neg[+\chi]\varphi$. For the right-to-left direction of (16), simply reverse the direction of the implications. It is also immediate from the truth definitions that (17) is valid. For (18) and (19), $[+\chi]$ and $\downarrow x.$ commute and $[+\chi]$ and $\Box^{\twoheadrightarrow}$ commute because the $+\chi$ operations do not change the assignment function g or the relation $\twoheadrightarrow$ from the initial model $\mathcal{M}$ to the new model $\mathcal{M}^{+\chi}$.

For (20), the lhs expresses that after context change by $\uparrow \chi$, all worlds that are more relevant than the current world w according to $\preceq^{\uparrow\chi}_{g(x)}$ satisfy φ:

$$\{v \in W \mid v \prec^{\uparrow\chi}_{g(x)} w\} \subseteq [\![\varphi]\!]^{\mathcal{M}^{\uparrow\chi}}. \tag{24}$$

Case 1: $\{v \in W \mid v \prec^{\uparrow\chi}_{g(x)} w\} = \emptyset$. This implies (24) and is equivalent to

$$w \in \mathrm{Min}_{\preceq^{\uparrow\chi}_{g(x)}}(W). \tag{25}$$

By Definition 5 for $\uparrow$, (25) holds iff either

$$w \in \mathrm{Min}_{\preceq_{g(x)}}(W), \tag{26}$$

which is equivalent to $\mathcal{M}, g, w \vDash \Box^{\succ_x}\bot$, or else

$$w \in \mathrm{Min}_{\preceq_{g(x)}}([\![\chi]\!]^{\mathcal{M}}), \tag{27}$$

which is equivalent to $\mathcal{M}, g, w \vDash \chi \wedge \Box^{\succ_x}\neg\chi$. This accounts for the first two disjuncts on the rhs of (20).

Case 2: $\{v \in W \mid v \prec^{\uparrow\chi}_{g(x)} w\} \neq \emptyset$. In this case, by Definition 5 for $\uparrow$,

$$\{v \in W \mid v \prec^{\uparrow\chi}_{g(x)} w\} = \{v \in W \mid v \prec_{g(x)} w\} \cup \mathrm{Min}_{\preceq_{g(x)}}([\![\chi]\!]^{\mathcal{M}}). \tag{28}$$

Hence (24) requires that

$$\{v \in W \mid v \prec_{g(x)} w\} \subseteq [\![\varphi]\!]^{\mathcal{M}^{\uparrow\chi}} = [\![[\uparrow \chi]\varphi]\!]^{\mathcal{M}}, \tag{29}$$

which is equivalent to $\mathcal{M}, g, w \vDash \Box^{\succ_x}[\uparrow \chi]\varphi$, and

$$\mathrm{Min}_{\preceq_{g(x)}}([\![\chi]\!]^{\mathcal{M}}) \subseteq [\![\varphi]\!]^{\mathcal{M}^{\uparrow\chi}} = [\![[\uparrow \chi]\varphi]\!]^{\mathcal{M}}, \tag{30}$$

which is equivalent to $\mathcal{M}, g, w \vDash \Box^x((\chi \wedge \Box^{\succ_x}\neg\chi) \rightarrow [\uparrow \chi]\varphi)$. The conjunction of $\Box^{\succ_x}[\uparrow \chi]\varphi$ and $\Box^x((\chi \wedge \Box^{\succ_x}\neg\chi) \rightarrow [\uparrow \chi]\varphi)$ is equivalent to

$$\Box^{\succ_x}[\uparrow \chi]\varphi \wedge \Box^{\preceq_x}((\chi \wedge \Box^{\succ_x}\neg\chi) \rightarrow [\uparrow \chi]\varphi), \tag{31}$$

which is the last disjunct on the rhs of (20).

For (21), what the lhs expresses about the current world w is

$$\{v \in W \mid w \preceq^{\uparrow\chi}_{g(x)} v\} \subseteq [\![\varphi]\!]^{\mathcal{M}^{\uparrow\chi}}. \tag{32}$$

Case 1: $\{v \in W \mid w \preceq^{\uparrow\chi}_{g(x)} v\} = W_{g(x)}$. This is equivalent to (25), which explains the first conjunct of the first disjunct on the rhs of (21). In this case, (32) requires that

$$W_{g(x)} \subseteq [\![\varphi]\!]^{\mathcal{M}^{\uparrow\chi}} = [\![[\uparrow \chi]\varphi]\!]^{\mathcal{M}}, \tag{33}$$

which is equivalent to $\mathcal{M}, g, w \vDash \Box^x[\uparrow \chi]\varphi$. This accounts for the second conjunct of the first disjunct on the rhs of (21).

Case 2: $\{v \in W \mid w \preceq^{\uparrow\chi}_{g(x)} v\} \neq W_{g(x)}$. In this case, by Definition 5 for $\uparrow$,

$$\{v \in W \mid w \preceq^{\uparrow\chi}_{g(x)} v\} = \{v \in W \mid w \preceq_{g(x)} v\} \setminus \mathrm{Min}_{\preceq_{g(x)}}([\![\chi]\!]^{\mathcal{M}}). \tag{34}$$

Hence (32) requires that

$$\{v \in W \mid w \preceq_{g(x)} v\} \setminus \mathrm{Min}_{\preceq_{g(x)}}([\![\chi]\!]^{\mathcal{M}}) \subseteq [\![\varphi]\!]^{\mathcal{M}^{\uparrow\chi}} = [\![[\uparrow \chi]\varphi]\!]^{\mathcal{M}}, \tag{35}$$

which is equivalent to $\mathcal{M}, g, w \vDash \Box^{\preceq_x}((\chi \wedge \Box^{\succ_x}\neg\chi) \vee [\uparrow \chi]\varphi)$. This explains the second disjunct on the rhs of (21). The arguments for (22) - (23) are similar. □

Given Propositions 1 and 2, if we combine the epistemic and RA languages and interpret $K\varphi$ according to D-semantics (a similar point holds for L), then we can write a reduction axiom for context change and knowledge as follows:

$$[+\chi]K\psi \leftrightarrow \downarrow x.\Box^{\rightarrow}(\neg[+\chi]\sigma_d(\psi) \rightarrow \neg\alpha), \tag{36}$$

where α is the rhs of (20) if $+$ is $\uparrow$ (resp. of (22) if $+$ is $\nwarrow$) with $\varphi := \sigma_d(\psi)$. Here we have used the fact that $\Diamond^{\succ_x}\neg\sigma_d(\psi)$ is equivalent to $\neg\Box^{\succ_x}\sigma_d(\psi)$, and $[+\chi]\neg\Box^{\succ_x}\sigma_d(\psi)$ reduces to $\neg[+\chi]\Box^{\succ_x}\sigma_d(\psi)$, which in turn reduces to $\neg\alpha$.

An important technical and conceptual issue raised by a result like Proposition 2 concerns the distinction between *valid* and *schematically valid* principles of context change. Where a principle is schematically valid just in case all of its substitution instances are valid [2, §3.12], the valid reduction principle $[+\chi]\, p \leftrightarrow p$ is clearly not schematically valid. Observe that $[+\chi]\, Kp \leftrightarrow Kp$ is not valid; if it were, there would be no epistemic dynamics. A more interesting example is the valid principle $\neg Kp \rightarrow [+\chi]\neg Kp$, which holds for our operations that make the context more epistemically "demanding." Observe that $\neg K\psi \rightarrow [+\chi]\neg K\psi$ is not valid for all ψ; it is possible to count as having some knowledge after the context becomes more demanding that one did not count as having before. How can this be? The answer is that this new knowledge may be knowledge of *ignorance*.[11] This can be seen by substituting $\neg Kp$ for ψ and either trying out model changes

[11] This is easiest to understand in a multi-agent setting. (Note that all of our definitions easily generalize to the multi-agent case where the modal operators in our language

or using (36) to reduce $\neg K\neg Kp \to [+\neg p]\neg K\neg Kp$ to a static principle that can be seen to be invalid. These observations raise the question, which we leave open, of what is the complete set of schematically valid principles of context change.

We leave as another open problem the task of finding an axiomatization of the theory of RA models in the static RA language (or some static extension thereof), which together with the reduction axioms of Proposition 2 would give an axiomatization of the theory of RA models in the dynamic RA language to go alongside the axiomatization in the epistemic language given by Theorem 1.

4 Conclusion

We have touched on two sides of RA theory, static (§2) and dynamic (§3), setting up a formal framework to study both. The range of results obtainable in this framework and its extensions, as well as their philosophical repercussions, are explored in [21]. On the dynamic side, having formally defined context change operations, we can see more clearly the systematic relations between theories that accept closure failures (Response 1 in §1) and theories that try to explain away closure failures in terms of context change (Response 2 in §1). On the static side, by using models like our RA models, we can characterize the epistemic closure properties not only for RA theories, but also for a family of "subjunctivist" theories that posit counterfactual conditions on knowledge, as well as the relations between these theories [20]. Moreover, these formalizations do not only help us to clarify the landscape of standard theories. They can also help us to see beyond the standard theories to new and improved pictures of knowledge.

Acknowledgements. I wish to thank Johan van Benthem, Tomohiro Hoshi, Thomas Icard, Krista Lawlor, and Eric Pacuit for helpful discussions and the anonymous reviewers for helpful comments on this paper.

References

1. Carlos Areces and Balder ten Cate. Hybrid Logics. In Patrick Blackburn, Johan van Benthem, and Frank Wolter, editors, *Handbook of Modal Logic*, pages 821–868. Elsevier, 2007.
2. Johan van Benthem. *Logical Dynamics of Information and Interaction*. Cambridge University Press, 2011.
3. Johan van Benthem, Patrick Girard, and Olivier Roy. Everything else being equal: A modal logic for *Ceteris Paribus* preferences. *Journal of Philosophical Logic*, 38:83–125, 2009.

and relations in our models are indexed for different agents.) Taking $\psi := \neg K_j p$, suppose agent i believes of agent j that $\neg K_j p$, but i does not know $\neg K_j p$, as i has not eliminated some relevant $K_j p$-worlds. If the context changes in such a way that j no longer counts as knowing p under any circumstances, then relative to this new context, i can count as knowing $\neg K_j p$. We can no longer fault i for not having eliminated some relevant $K_j p$-worlds if there are none relative to the current context.

4. Johan van Benthem and Fenrong Liu. Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logics*, 17(2):157–182, 2007.
5. Oliver Board. Dynamic interactive epistemology. *Games and Economic Behavior*, 49:49–80, 2004.
6. Brian F. Chellas. *Modal Logic: An Introduction*. Cambridge University Press, 1980.
7. Stewart Cohen. How to be a Fallibilist. *Philosophical Perspectives*, 2:91–123, 1988.
8. Stewart Cohen. Basic Knowledge and the Problem of Easy Knowledge. *Philosophy and Phenomenological Research*, 65(2):309–329, 2002.
9. Keith DeRose. Solving the Skeptical Problem. *The Philosophical Review*, 104(1):1–52, 1995.
10. Keith DeRose. *The Case for Contextualism*. Oxford University Press, 2009.
11. Hans van Ditmarsch, Wiebe van der Hoek, and Barteld Kooi. *Dynamic Epistemic Logic*. Springer, 2008.
12. Fred Dretske. Epistemic Operators. *The Journal of Philosophy*, 67(24):1007–1023, 1970.
13. Fred Dretske. The Pragmatic Dimension of Knowledge. *Philosophical Studies*, 40:363–378, 1981.
14. Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. *Reasoning about Knowledge*. MIT Press, 1995.
15. Alvin I. Goldman. Discrimination and Perceptual Knowledge. *The Journal of Philosophy*, 73(20):771–791, 1976.
16. Joseph Y. Halpern and Riccardo Pucella. Dealing with Logical Omniscience: Expressiveness and Pragmatics. *Artificial Intelligence*, 175:220–235, 2011.
17. Marc Heller. Relevant Alternatives and Closure. *Australasian Journal of Philosophy*, 77(2):196–208, 1999.
18. Mark Heller. Relevant Alternatives. *Philosophical Studies*, 55:23–40, 1989.
19. Jaakko Hintikka. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. College Publications, 2005.
20. Wesley H. Holliday. Epistemic Closure and Epistemic Logic I: Relevant Alternatives and Subjunctivism. Manuscript, 2012.
21. Wesley H. Holliday. *Knowing What Follows: Epistemic Closure and Epistemic Logic*. PhD thesis, Stanford University, 2012.
22. David Lewis. Completeness and Decidability of Three Logics of Counterfactual Conditionals. *Theoria*, 37(1):74–85, 1971.
23. David Lewis. *Counterfactuals*. Oxford: Blackwell, 1973.
24. David Lewis. Scorekeeping in a Language Game. *Journal of Philosophical Logic*, 8:339–359, 1979.
25. David Lewis. Elusive Knowledge. *Australasian Journal of Philosophy*, 74(4):549–567, 1996.
26. Robert Nozick. *Philosophical Explanations*. Harvard University Press, 1981.
27. Robert Stalnaker. The Problem of Logical Omniscience I. *Synthese*, 89:425–440, 1991.