Opaque Updates

August 10, 2020

Abstract

If updating with E has the same result across all epistemically possible worlds, then the agent has no uncertainty as to the behavior of the update, and we may call it a transparent update. If an agent is uncertain about the behavior of an update, we may call it opaque. In order to model the uncertainty an agent has about the result of an update, the same update must behave differently across different possible worlds. In this paper, I study opaque updates using a simple system of dynamic epistemic logic suitably modified for that purpose. The paper highlights the connection between opaque updates and the dynamic-epistemic principles Perfect-Recall and No-Miracles. I argue that opaque updates are central to contemporary discussions in epistemology, in particular to externalist theories of knowledge and to the related problem of epistemic bootstrapping, or easy knowledge. Opaque updates allow us to explicitly investigate a dynamic (or diachronic) form of uncertainty, using simple and precise logical tools.

Keywords: dynamic epistemic logic, externalism, basic knowledge, bootstrapping, introspection, Perfect-Recall, No-Miracles, Bayesian update, opacity.

1 Introduction

There is a widespread notion of *update* in formal epistemology that can be semantically summarized in Figure 1.

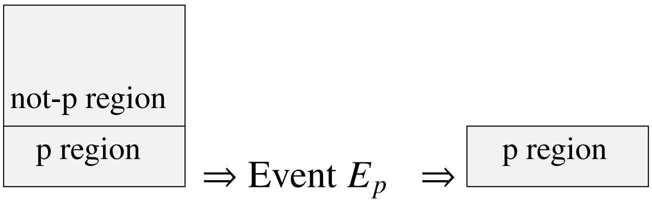

Figure 1: Update as a world insensitive function.

We have some prior, or initial, model on the left-hand of the Figure, containing both *p* and not-*p* worlds, representing a situation in which an agent is uncertain whether *p* is the case. The event *E* of receiving the information that *p* results in a new model in which the not-*p* worlds are eliminated (the model to the right), a model representing a situation in which there is no uncertainty as to *p*. This simple sketch of updating is at the basis of many systems that represent information change, including the Stalnakerian notion of assertion, Dynamic Semantics, Bayesian updating, and various dynamic epistemic logics.

Of course, each implementation of this basic skeleton idea is different, but here I want to point our attention to an assumption that can be detected even at this level of abstraction. The picture in Figure 1 portrays the event of learning *p* as a transition from one model to the other, a model transformer. In other words, the event of receiving the information that *p* is understood as a function (sometimes partial) from epistemic models to epistemic models. Since such a function is assumed to send us from one *model* to another, it is *insensitive* to the world of evaluation. In other words, the function behaves the same at every possible world of the prior model—at each world, the function sends us to the same posterior model. Put differently again, we don't need to know which world is considered actual in order to compute the model which results from an update with *p*. At the same time, the picture in Figure 1 builds on the idea that when something is the same in all possible worlds, there is no uncertainty about it. Putting these two threads together, since the picture in Figure 1 assumes that the update behaves the same in each possible world, and since certainty is assumed to be agreement across all possible worlds, we essentially assume that *the agent has no uncertainty as to the behavior of the update*. Thus, the update in Figure 1 is in some sense *transparent* to the agent.

In this paper, I analyze updates that are not transparent, but *opaque*. If we want to represent the uncertainty the agent has about the effect of an update, it should be possible to have situations in which the same update behaves differently in different epistemically possible worlds. Such situations are not just meant to generalize the notion of update for purely technical reasons. The second main theme of my paper is that modeling opaque updates is, as I will argue, quite relevant to various debates in contemporary epistemology, and in particular, to the broad position of *externalism*.

For instance, a reliabilist about knowledge argues that the effect of the same learning event can be different in a world in which the source of information is reliable as opposed to a world in which the source is unreliable, while assuming that the agent does not know if the source is in fact reliable. Consider a simple scenario: you look at a clock tower that has, in fact, a reliable clock mechanism. You don't know, however, that the clock mechanism is reliable. According to reliabilists (and many other externalists about knowledge and evidence), it is possible for you to come to know the time by looking at the clock, even though you don't know the clock is reliable. Thus, prior to looking at the clock, you cannot know whether looking at the clock will result in a situation in which you know the time (if the clock is in fact reliable) or in a situation in which you don't know the time (if it is unreliable). Externalists think of the event of looking at the clock as *opaque*: the agent is uncertain as to the epistemic result of the event.

In this paper, we will see how to model such situations in a simple possible-worlds framework. The formal idea is to build an update U s.t. U sends us to different updated models in different worlds of the prior model. In world w of the prior model in which r is true (the clock is reliable), U will send us to an updated model in which the agent knows the time. In world u of the prior model in which $\neg r$ is true (the clock is unreliable), U will send us to a different updated model in which the agent does not know the time. Since both worlds u and w are initially open to the agent, the agent does not know whether the update with U will result in knowledge of the time or not. We will also explore a backwards form of diachronic uncertainty involving the posterior model: opaque updates can be such that the agent does not know whether it is update U that brought them to their posterior epistemic state, or U' (an update distinct from U).

Externalism about knowledge is roughly the idea that factors external to the epistemic agent can determine the difference between having knowledge and having mere true belief, even if the agent cannot notice those factors from their own (internal) perspective. Reliabilism is just one example. In this paper, I show how to formalize the externalist intuition when it comes to *epistemic change*. It is possible to construct epistemic updates whose results depend on external factors that the agent is ignorant about. The updates that capture the externalist intuition are opaque updates.

The epistemic logic literature already contains influential endeavors to formalize externalist intuitions. The important works of Rott (2004), Stalnaker (2006) and Baltag and Smets (2008) show how to obtain an externalist conception of knowledge as belief that is stable under revision with true information. However, these approaches do not focus on at least two issue that have become increasingly relevant in the contemporary externalist literature. First, since the above mentioned approaches rely on a 'sphere system' semantics to model belief revision, it follows that the resulting conception of knowledge is positively introspective, i.e. validates the KK principle which states that $K\varphi \to KK\varphi$. Since the vast majority of externalists in epistemology take the *rejection* of the KK principle as essential to externalism, they will reject these formal approaches.[1] Second, the above mentioned formulations seek to analyze knowledge in terms (of some properties) of belief. A prominent camp within externalism, the 'knowledge-first' one, dismisses the

---

[1] See Okasha (2013) for a survey on the connection between externalism and the rejection of KK.

project of analyzing knowledge in terms of belief.[2] Therefore, knowledge-first externalists will not accept these approaches. The formulation I propose here can shed new light on these issues: the logic I present for opaque updates is compatible both with the acceptance of the KK principle and with its rejection. Moreover, the logic does not make any assumptions about the relations between knowledge and belief (a belief modality can of course be added to the logic, but I will not explore this here). The model I offer aims to remain as neutral as possible in its epistemological assumptions.

Timothy Williamson has also offered models of epistemic logic for studying externalist conceptions of knowledge (2000, 2013, 2014). In these models, the agent's *lack* of positive introspection plays a central role. However, Williamson's models are completely static, remaining silent with respect to the question how externalism construes *change of knowledge*. The dynamic formulation I offer here is meant to bridge this gap.

Within formal epistemology, we are used to model the information the agent has as a set of possible worlds. *Static* epistemic logic has taught us that assuming that such a set is constant across possible worlds amounts to assuming that the agent has *full static* introspection. Philosophical work has concluded that such assumption is highly debatable. Analogously, dynamic epistemic logic teaches us, as I will argue, that assuming that the result of an update is constant across possible worlds amounts to assuming that the agent has *full dynamic* introspection, or full dynamic transparency. Is such assumption justified?

The aim of this paper is to introduce, analyze and offer a simple working model for an epistemic-logic phenomenon: opaque updates. Rather than advancing an entirely novel contribution to the logical literature, or offering a philosophical argument for or against externalism, my goal is to *bridge* logical and epistemological bodies of work. I show how the right application of existing logical tools can be used to offer a fresh approach to well-known problems in epistemology, problems that can benefit from an accessible formal model.

In Section 2 of this paper, I show how to expand a simple dynamic epistemic logic such that it will be able to accommodate opaque updates. Along the way, I highlight and discuss important epistemological assumptions that are built into basic dynamic epistemic logics. The extension uses a familiar combination of propositional dynamic logic with dynamic epistemic logic. Such an extension allows us to violate the basic update axioms of dynamic epistemic logic, the No-Miracles and Perfect-Recall axioms. The combination of these two axioms can be seen as the syntactic analog of the semantic idea from Figure 1. In Section 3, I argue for the relevance of opaque updates (and the rejection of the No-Miracles and Perfect-Recall axioms) to what is known as *basic knowledge* theories within

---

[2] See Williamson (2000), Nagel (2013), and the collection of papers in Carter et al. (2017).

epistemology, and to the related discussion about the problem of epistemic bootstrapping. Section 4 discusses and compares the phenomenon of opaque updates in other logical frameworks. Section 5 concludes. Although the technical details of opaque updates are presented for a particular version of dynamic epistemic logic, the broader aim of this paper is to introduce notions and principles that go beyond any particular system.

2 Dynamic epistemic logic with opaque updates

We will use the single-agent *Public Announcement Logic* (PAL) as our base dynamic epistemic logic, which will then be extended to accommodate opaque updates that violate the axioms of PAL. Before we get there, I start with a brief recap of standard PAL. Readers familiar with the basics of PAL can jump to Section 2.2.

2.1 Public Announcement Logic

Public announcement logic is an extension of static epistemic logic with simple epistemic events, announcements, that transmit true information reliably and publicly (to all agents) (Baltag and Renne 2016). For our purposes, we can think of single-agent PAL as perhaps the simplest logical system that follows the semantic idea presented in Figure 1. As an extension of epistemic logic, we have a modal propositional operator K, representing the propositional knowledge of the agent. We further have an update operator $[!\varphi]$ for every φ of the language, s.t. $[!\varphi]\psi$ is read "as a result of the announcement of (or update with) φ, ψ is the case." Diamond duals of the modal operators are defined as usual: $\hat{K}\varphi$ is defined as $\neg K \neg\varphi$ and $\langle !\varphi\rangle\psi$ is defined as $\neg[!\varphi]\neg\psi$.

A formula φ of PAL is evaluated over a Kripke model $M = (W, R, V)$, where W is a non-empty set of possible worlds, R is an epistemic indistinguishability relation that is assumed to be reflexive,[3] and V is a valuation function, mapping every atomic formula to a subset of W. Formulas are evaluated with respect to a pair (M, w) of a Kripke model and a specific point $w \in W$ as standard in modal logic. In particular, the formula $K\varphi$ is true at a possible world w iff all worlds u that are accessible to w via R are worlds in which φ is true. Formally:

$M, w \models K\varphi \quad \Leftrightarrow \quad \forall u : wRu, M, u \models \varphi.$

The semantic idea behind the $[!\varphi]$ operator is exactly the one from Figure 1: updating the Kripke model with φ results in a model in which all the not φ worlds are eliminated. Since PAL updates are assumed to be veridical, $[!\varphi]\psi$ is taken to be vacuously true in a world in which φ is false, while $\langle !\varphi\rangle\psi$ is taken to be false in such world (this sums up the

---

[3] We are not assuming that R is transitive or Euclidean.

difference between the box and diamond versions of the update operator). More formally, we have the following semantic clause for the update operator:
$M, w \models [!\varphi]\psi \quad \Leftrightarrow \quad M, w \models \varphi \ \textit{implies} \ M_\varphi, w \models \psi.$

The antecedent of the right-hand side of this condition guarantees that when we try to update with φ is a world in which φ is false, $[!\varphi]\psi$ is vacuously true. The consequent states that ψ must be true in the model which results from updating M with φ, which is denoted as M_φ and defined as $(W_\varphi, R_\varphi, V_\varphi)$, where W_φ is just $W \cap \{w \in W \mid M, w \models \varphi\}$ and R_φ and V_φ are the restrictions of R and V with respect to W_φ. With relation to the discussion around Figure 1, we can think of $[\cdot]_\varphi$ as a function sending us from Kripke models to Kripke models, s.t. $[M]_\varphi = M_\varphi$. Note that the value of such function is indeed insensitive to the world of evaluation; we don't need an actual world to compute $[M]_\varphi$ from M.[4]

For a simple example, consider Ann, who looks at a tower clock and learns that the time is 12:00 (let that be proposition p). The models in Figure 2 depict the PAL update with such p.

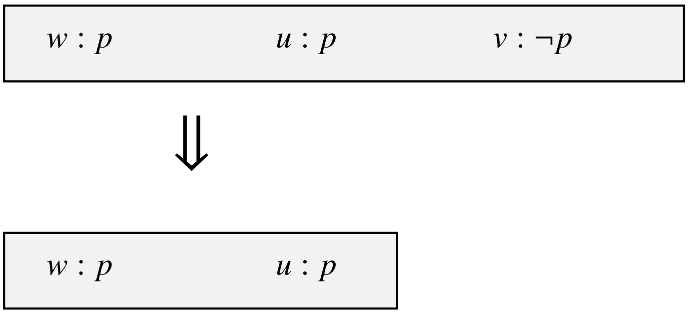

Figure 2: A standard PAL update

Figure 2 depicts a transition from the prior model M at the top to the posterior model M_p at the bottom. The epistemic R relation is depicted with the grey box. Note that since world v does not satisfy p in the initial model, it is eliminated from the posterior model M_p. In order to compute M_p we don't need to know which world is considered actual. The PAL model transition in Figure 2 is just one example of the semantic idea from Figure 1.

2.1.1 Reduction axioms for PAL

A popular way to axiomatize PAL is via the following set of *reduction axioms*[5]

---

[4] In the current presentation of PAL, the world of evaluation does determine whether the partial function $[\cdot]_\varphi$ is defined or not. This is the only sense we can say that the partial function $[\cdot]_\varphi$ is sensitive to the world of evaluation.

[5] Note that for each such axiom, the complexity of β in the sub-expression $[!\alpha]\beta$ is reduced in the right-hand side as compared to the left-hand side (although the overall complexity of the right-hand side expression increases). See Baltag and Renne (2016) for more information.

1. $[!\varphi]p \leftrightarrow (\varphi \to p)$ atomic reduction
2. $[!\varphi]\neg\psi \leftrightarrow (\varphi \to \neg[!\varphi]\psi)$ negation reduction
3. $[!\varphi](\chi \wedge \psi) \leftrightarrow ([!\varphi]\chi \wedge [!\varphi]\psi)$ conjunction reduction
4. $[!\varphi]K\psi \leftrightarrow (\varphi \to K[!\varphi]\psi)$ knowledge reduction
5. $[!\varphi][!\psi]\chi \leftrightarrow [!\varphi \wedge [!\varphi]\psi]\chi$ announcement reduction

For every formula $[!\varphi]\psi$ of PAL, a repeated application of the reduction axioms results in an equivalent expression ψ^* in which the update operator has been eliminated. In other words, every posterior epistemic state can be syntactically manipulated into an expression about the initial epistemic state. This is similar to the fact that every posterior probability function can be syntactically manipulated into a conditional prior probability function in the Bayesian framework. The common idea is that since the prior epistemic state *determines* every posterior state, every posterior state can be reduced back to the prior state.

2.2 Forest models for PAL

While the standard semantics and axiomatization of PAL is quite straightforward, it is not incredibly helpful in highlighting the transparency property of PAL updates. The world elimination semantics do not leave much wiggle room for modifications with the effects of updates, and the reduction axioms do not illuminate any particular principle about the agent's knowledge of the effects of updates, nor do they directly correspond to any semantic frame condition, like many other modal axioms. Still, for any Boolean formula β it is easy to see (either syntactically or semantically) that $K[!\beta]K\beta$ is valid in standard PAL: the PAL agent has the prior knowledge that any update with β results in knowledge that β – there is no uncertainty as to the reliability of (Boolean) updates, they always work.[6]

In order to break the rules of PAL, we will start by presenting an alternative but equivalent semantics for PAL, and a set of axioms that fits nicely with the alternative semantics. The alternative system can be then easily tweaked to construct opaque updates that violate the principles of PAL. The alternative semantics is well-known in the dynamic epistemic logic literature; the novelty of this paper lies in the way we expand that semantics to analyze opaque updates.

---

[6] The formula $K[!\varphi]K\varphi$ is not valid in general in PAL, however. This is because of Moorean sentences like $p \wedge \neg Kp$ which can change truth value after the announcement. Nevertheless, the PAL agent has no uncertainty as to the reliability of their information sources.

The idea of the alternative PAL semantics is to treat the epistemic result of an update not as a new epistemic model, but as a different part of one big model that contains all its updated models as submodels. Very informally, the idea is to make the meta-theoretic arrow in Figure 1 into a regular object-level relation of some Kripke model. On such semantics, the interpretation of the modal expression $[!\varphi]\psi$ is the familiar truth of ψ in all the relevant worlds. Such semantics have been studied extensively in van Benthem et al. (2009) and Wang and Cao (2013).

In order to evaluate $[!\varphi]$ as a regular modal operator, we need to be able to expand a given Kripke model with a $\rightarrow_\varphi$ relation for every formula φ and add new possible worlds for the result of updates. If w is part of the initial epistemic model, and φ is a sentence true in w, then we add the world (w, φ) to the model. World (w, φ), which represents the situation resulting from learning φ in w, has the same atomic valuation as w, and it is connected to w via the $\rightarrow_\varphi$ relation. Such an expanded model is known as the *forest* model of the original model. Here is one way to construct such a forest model from a given epistemic model M (modified from Yap and Hoshi (2009), Aucher and Herzig 2010). The definition is followed by an informal explanation.

PAL Forest Model: Given a reflexive Kripke model $M = (W, R, V)$ we define its PAL forest $Forest(M) = (F(W), F(R), \rightarrow_\varphi F(V))$ s.t.
–$F(W) = \bigcup_n W^n$
–$F(R) = \bigcup_n R^n$
– $\rightarrow_\varphi = \bigcup_n \rightarrow^n_\varphi$
–$F(V) = \bigcup_n V^n$,
where $M^n = (W^n, R^n, \rightarrow^n_\varphi V^n)$ is defined inductively as follows:
– $M^0 = M$, where $\rightarrow^0_\varphi = \emptyset$.
– $M^{n+1} =$

- $W^{n+1} = W^n \cup \{(w, \varphi) \ : \ w \in W^n, \varphi \in \mathcal{L}_{PAL} \ \& \ M^n, w \models \varphi\}$
- $R^{n+1} = R^n \cup \{(w, \varphi), (v, \psi) \ : \ wR^n v \ \& \ \varphi = \psi\}$
- $\rightarrow^{n+1}_\varphi = \rightarrow^n_\varphi \cup \{w, (w, \varphi) \ : \ w \in W^n\}$
- $V(p)^{n+1} = V(p)^n \cup \{(w, \varphi) \ : \ (w, \varphi) \in W^{n+1} \ \& \ w \in V(p)^n\}$

Here is an informal description of the construction in the definition. We want to create a relation for every possible PAL update φ. $Forest(M)$ is the union of all M^n models, where M^0 represents the original model we start with, M^1 represents the model after one update with φ, M^2 after two updates with φ and so on. Starting at a world w in M^0, if φ is false in that world, then, since PAL updates are veridical, we don't connect any $\rightarrow_\varphi$ to that world w. If φ is true at w, we add a new world to the forest model, the world (w, φ), which is part of W^1. That world has the same atomic valuation as w. We then connect w to (w, φ) with the $\rightarrow_\varphi$ relation. We extend the epistemic R relation to model M^1 s.t. if w was accessible to u in model M^0, then (w, φ) will be accessible to (u, φ) in model M^1 (assuming that the two worlds exist in M^1). Since we can repeatedly update with φ again and again, the forest construction is infinite.

Figure 3 contains a simple example of a forest model.

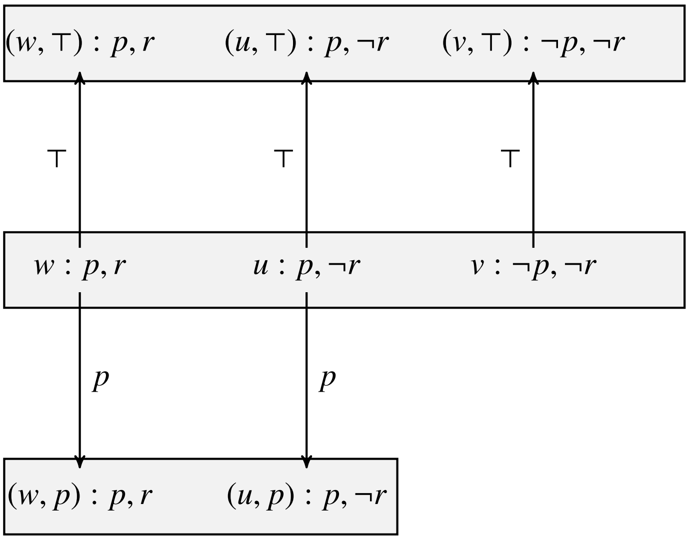

Figure 3: A (partial) forest construction

The model in Figure 3 (partially) depicts a forest construction, where the middle box is the initial Kripke model from which we extend to a forest construction. The epistemic R relation is an equivalence relation represented by the grey boxes (worlds in the same grey box are connected with the R relation). The black arrows represent the update relations $\rightarrow_p$ and $\rightarrow_\top$. The $\rightarrow_p$ relation depicts the update with the propositional letter p, the result of that update can be seen in the lower submodel. The $\rightarrow_\top$ depicts the result of updating with a tautology, which has no effect and results in a copy of the original model, represented in the upper submodel.[7]

The alternative evaluation (denoted $\models_a$) of PAL updates is very simple on forest models: we just follow the $\rightarrow_\varphi$. Considering Figure 3, we can see that $M, w \models_a [!p](Kp \land \neg Kr)$ by following the single p arrow from w and noting that $M, (w, p) \models_a Kp \land \neg Kr$; after the update with p, the agent knows p but does not know r.

---

[7] The model does not depict other updates, like the update with r or $\neg p$, nor does it depicts iterated updates with p or $\top$, although it is clear that such updates have reached a fixed point.

2.3 An alternative axiomatization of PAL

The $\models_a$ evaluation is equivalent to the standard PAL evaluation from Section 1.1 on Forest constructions (Wang and Cao 2013), but, as we will soon see, the alternative semantics on the forest construction can be quite illuminating when it comes to opaque updates. With the alternative semantics, Wang and Cao (2013) have shown that the following set of axioms axiomatizes PAL as well:[8]

6. $(p \to [!\varphi]p) \land (\neg p \to [!\varphi]\neg p)$ Atomic Invariance
7. $\langle !\varphi\rangle\psi \leftrightarrow \varphi \land [!\varphi]\psi$ Partial Function
8. $\langle !\varphi\rangle K\psi \to K[!\varphi]\psi$ No-Miracles
9. $K[!\varphi]\psi \to [!\varphi]K\psi$ Perfect-Recall

This way of presenting the axioms of PAL is valuable both because each axiom corresponds to a clear frame condition on forest models and because each axiom (but especially 7.-9.) represents *a non-trivial, debatable, epistemological commitment.*

The atomic invariance axiom states that epistemic updates do not change the non-epistemic facts in the world. Semantically, it corresponds to the conditions that if $x \to_\varphi y$ then x and y agree on every atomic formula. The axiom defines the events we analyze as *epistemic* events, rather than *ontic* events that change the non-epistemic facts of the world.

The Partial Function axiom essentially states that updates are *deterministic*, that given a particular situation, there is a fact of the matter as to how the update affects the agent (even if the agent does not know that). For a recent discussion about deterministic updates in a Bayesian context, see Pettigrew (2019). Semantically, the axiom corresponds to the fact that each world in the forest model has *at most* one $\to_\varphi$ coming out of it (for each φ). In my epistemological application of opaque updates (Section 3), I assume that updates are deterministic.

2.3.1 The No-Miracles and Perfect-Recall principles

The last two axioms, No-Miracles (NM) and Perfect-Recall (PR), are of the most importance when it comes to understanding the difference between transparent and opaque updates.[9] One way of thinking about NM and PR is as *dynamic introspection principles*,

---

[8] Together with a distribution (K) axiom and a necessitation rule for the operator $[!\varphi]$.

[9] The No-Miracles principle is not related to the No-Miracles argument from philosophy of science. A better name for NM might be No-Surprises, since, informally, the principle expresses the idea that the agent is never surprised by the result of an update. Here we follow the epistemic logic literature and stick with the name NM. For more on NM and PR in epistemic logic, see van Benthem et al. (2009), van Benthem

which contrasts them to the well-known, and well debated, static introspection principles like $K\varphi \rightarrow KK\varphi$ and $\neg K\varphi \rightarrow K\neg K\varphi$ (positive and negative introspection, respectively). Static introspection principles involve what we know about our own mental states; syntactically, such principles involve the scoping of the K operator over other instances of the K operator. Analogously, I suggest we call NM and PR dynamic introspection principles because they involve what we know about our own epistemic events. Syntactically, dynamic introspection principles involve the scoping of the K operator over the update operator.[10]

When it comes to ignorance about updates, two independent types of questions arise: future (or forward) directed and past (or backward) directed. The future directed question asks, given a particular event, how is that event going to epistemically affect me: *where am I going from here?* The NM principle ($\langle !\varphi\rangle K\psi \rightarrow K[!\varphi]\psi$) answers this question in the following way. It states that if the update with φ actually results in the agent being in a position to know ψ ($\langle !\varphi\rangle K\psi$), then the agent has the *prior* knowledge that φ updates result in ψ being the case ($K[!\varphi]\psi$). Taking NM as an axiom amounts to assuming that the agent always has the ability to correctly predict the effect of updates, that the agent is never ignorant as to the result of the update. This is one part of what it means for an update to be transparent to the agent.

Past directed ignorance about updates can be summarized in the question: given my current epistemic position, what update exactly has brought me to this position? *How did I get here?* The PR principle ($K[!\varphi]\psi \rightarrow [!\varphi]K\psi$) offers an answer. PR states that if the agent has the prior knowledge that the update with φ results in a ψ state ($K[!\varphi]\psi$), then as a result of the $[!\varphi]$ update, the agent is in a position to know ψ ($[!\varphi]K\psi$). Consider the negation of PR, stating $K[!\varphi]\psi \wedge \neg[!\varphi]K\psi$. It exemplifies opaqueness towards the update φ. Assume that ψ is the 'mark' of a φ update. The agent knows that ψ is the mark of a φ event, but they don't know ψ after the φ update. This implies that the agent does not know that the update was a φ update. As we will see, however, the failure of PR does *not* imply that the update failed to convey information.

Both NM and PR can be challenged on the grounds of the cognitive limitations of actual, non-idealized, agents. Rejecting PR as a way of modeling an agent's memory loss has been discussed within Bayesian epistemology, but I am not aware of analogous discussions about NM.[11] A different question is the compatibility of NM and PR with

---

(2012), Wang and Cao (2013); for the connection with game theory, see van Benthem (2014), van Benthem and Klein (2018).

[10] NM and PR are logically independent from axioms 4. and 5. of epistemic logic (static positive and negative introspection, respectively.) Any combination of dynamic and static introspection is therefore possible.

[11] See, e.g., Arenzious (2003) and Halpern (2004) for discussions about PR. Traces of the underlying commitments behind NM and PR can be detected in the Bayesian conditionalization scheme $P_H(E) =$

various theories in epistemology, regardless of cognitive limitations; the next section of this work is devoted to such issues.

NM and PR, as modal axioms, correspond to frame conditions on forest models. NM states that if wRt and $t \rightarrow_{\varphi} v$, and there is a world u s.t. $w \rightarrow_{\varphi} u$, then uRv. PR states that if $w \rightarrow_{\varphi} u$ and uRv, then there is a world t s.t. wRt and $t \rightarrow_{\varphi} v$ (Wang and Cao 2013). The two semantic conditions can be elegantly summarized in the following figure:

$$\begin{tikzcd}
u \arrow[r, "R"] & v & & u \arrow[r, "R"] & v & & u & v \\
w \arrow[u, "\varphi"] & & & w \arrow[u, "\varphi"] \arrow[r, "R"] & t \arrow[u, "\varphi"] & & w \arrow[u, "\varphi"] \arrow[r, "R"] & t \arrow[u, "\varphi"] \\
& & \mathrm{PR} \Rightarrow & & & \Leftarrow \mathrm{NM} & &
\end{tikzcd}$$

Figure 4: NM and PR as semantic conditions

Informally, NM and PR together imply that forest models are commutative in the following sense: every world you can get to by first going via the R relation and then by the φ relation, you also get to by first going via the φ relation and then the R relation, and vice versa. Consult the forest model in Figure 3 for an example.

2.4 Adding opaque updates to PAL

We are now in a position to see how to construct updates that break the NM and PR properties. The idea I am going to present is to *compose* new update relations from existing PAL update relations. Consider again Figure 3. If we had a tool to pick subsets

---

$P(H|E)$, where P_E stands for the posterior probability function after learning (with certainty) E. If we break down conditionalization into two directions for a specific numerical constant c, say $c = 1$, we get the commitments:

(i) $P_E(H) = 1 \;\Rightarrow\; P(H|E) = 1$

(ii) $P(H|E) = 1 \;\Rightarrow\; P_E(H) = 1$

(i) and (ii) roughly correspond to instances of NM and PR, respectively. (i) states that if the agent actually becomes certain in H after the update with E (in the sense of assigning probability 1), then the agent has the *prior* certainty that H is the case, given E. (ii) states that if the agent starts with the prior certainty in H given E, then that certainty is not lost once E is actually learned. However, unlike NM and PR in PAL, which connect two dynamic expressions, conditionalization connects a dynamic expression (the posterior state) with a static attitude (prior conditional probability).

of the union of the $\top$ and φ relations, we could easily create new relations (updates) that don't respect the commutative structure implied by NM and PR. Fortunately, there exists a modal tool that allows us to reason about composition of relations, known as propositional dynamic logic, or PDL. Combining PDL with dynamic epistemic logic is a rich field of study.[12] In what follows, I present a simple way to combine the alternative forest semantics of PAL with a fragment of PDL to get a logic that is sufficiently flexible to model opaque updates.

Unlike PAL, the logic of opaque updates denotes updates with expressions π that themselves are not always wffs of epistemic logic. Such $[\pi]$ operators represent updates that might be basic PAL updates or some composition of PAL updates. The language of the logic of opaque updates is defined inductively as follows:

$$\varphi := \ p \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid K\varphi \mid [\pi]\varphi$$

$$\pi := \ !\psi \mid ?\psi \mid \pi_1 ; \pi_2 \mid \pi_1 \cup \pi_2$$

where the formula ψ is of the language of epistemic logic (i.e. does not contain the update operator).[13] In PDL, π is called a program. $!\psi$ is considered in the logic of opaque updates as an atomic program (from which we compose more complicated programs). Every PAL update is an atomic program in the logic of opaque updates. For example, $!p$ is the program whose execution amounts to a PAL update of the formula p. $?\psi$ is a test program, which checks if ψ is true and aborts otherwise. $\pi_1;\pi_2$ is a composite program that sequentially executes program π_1 and then π_2. $\pi_1 \cup \pi_2$ is a composite program that non-deterministically executes π_1 or π_2. The syntax of PDL is useful for expressing familiar algorithmic expressions: the expression "if α, do π_1, otherwise do π_2" is written as $(?\alpha;\pi_1) \cup (?\neg\alpha;\pi_2)$. In the next section we will see epistemological examples of such expressions. The expression $[\pi]\varphi$ reads "after every execution of π, φ is the case." The expression $\langle\pi\rangle\varphi$ reads "there is an execution of π after which φ is the case."

In what follows, we evaluate formulas of the logic of opaque updates over forest models. Given some epistemic Kripke model M, we evaluate φ on $Forest(M)$. The semantic clause for the K operator is the same as in epistemic logic. The semantic clause of the expression $[\pi]\varphi$ is straightforward:

$M, w \models [\pi]\varphi \quad \Leftrightarrow \forall u : wR_\pi u, M, u \models \varphi$,

where the relation R_π is composed inductively in the following manner:

---

[12] See Troquard and Balbiani (2019) for a survey of PDL. See, e.g., van Ditmarsch (2000), van Benthem et al. (2006), Wang et al. (2009), Girard et al. (2012) for works connecting PDL and DEL.

[13] The restriction on ψ simplifies the technical discussion with no effect on the philosophical part. We exclude the Kleene star operator of full PDL from this system. For the study of PAL with a Kleene star for updates, see Miller and Moss (2005).

- $wR_{!\psi}u$ iff $w \to_{\psi} u$,
- $wR_{?\psi}w$ iff $Forest(M), w \models \psi$,
- $wR_{\pi_1;\pi_2}u$ iff there is a v s.t. $wR_{\pi_1}v$ and $vR_{\pi_2}u$,
- $wR_{\pi_1 \cup \pi_2}u$ iff either $wR_{\pi_1}u$ or $wR_{\pi_2}u$.

For a few examples, consider $Forest(M)$ from Figure 3 again. The relation $R_{!p\cup!\top}$ is just the union of the $\to_p$ and $\to_\top$ relations in that model, and the relation $R_{?(p\wedge r);!p}$ only consists of the pair of worlds $(w, (w, p))$, since, out of w, u, and v, only world w satisfies $(p \wedge r)$.

2.5 Reduction axioms for opaque updates

The logic of opaque updates is completely axiomatized by adding to the PAL axioms the following PDL axioms:

10. $[?\psi]\varphi \leftrightarrow (\psi \to \varphi)$
11. $[\pi_1; \pi_2]\varphi \leftrightarrow [\pi_1][\pi_2]\varphi$
12. $[\pi_1 \cup \pi_2]\varphi \leftrightarrow [\pi_1]\varphi \wedge [\pi_2]\varphi$

Note that axioms 10.-12. are reduction axioms, meaning that every formula ψ of the logic of opaque updates can be reduced to a formula ψ^* in the language of PAL (i.e., a formula that has no updates with ?, ; or $\cup$). For example, the formula

$$[(?r; !p) \cup (?\neg r; !\top)]Kp$$

can be equivalently written as (using 10.- 12.)

$$(r \to [!p]Kp) \wedge (\neg r \to [!\top]Kp),$$

which is a PAL formula, and any PAL formula, as we already discussed, can itself be reduced to a formula without any update operators. In fact, we can think of the equivalences in 10. to 12. as the definitions for the syntactic abbreviations ?, ; and $\cup$. This shows that the forest constructions (and the alternative semantics that follows it) which we have used are not strictly necessary: we can think of any program π in the logic of opaque updates as an abbreviation of a PAL expression, which can be evaluated with the standard PAL semantics.

Like in PAL (and Bayesian updating for that matter), in the logic of opaque updates, every formula expressing a posterior epistemic state can be rewritten as an expression without an update, describing the agent's prior state. The difference is that in the logic of opaque updates the truth value of the reduced sentence will not only depend on the epistemic situation, but also on non-epistemic factors (e.g., the truth of r in the above example). The prior situation still determines everything, but it must include more than just the prior *epistemic* situation.

Obviously, NM and PR will not be valid principles in a logic designed to model opaque updates. We can formulate the latter principles in our richer language as

13. $\langle\pi\rangle K\varphi \rightarrow K[\pi]\varphi$ No-Miracles

14. $K[\pi]\varphi \rightarrow [\pi]K\varphi$ Perfect-Recall

The semantic conditions of NM and PR, depicted in Figure 4, can also be generalized by replacing the φ relations with π relations. The next section presents concrete counterexamples to formulas 13. and 14., motivated by recent debates in contemporary epistemology.

3 Opaque updates in contemporary epistemology

This Section discusses a few applications of opaque updates in debates in epistemology. The applications center around issues concerning externalist conceptions of knowledge, broadly understood. The below Subsections are not meant to settle the philosophical discussions one way or the other, rather to show that opaque updates, and the NM and PR principles, are useful and relevant conceptual tools in such discussions.

3.1 Basic knowledge theories

Following Cohen (2002), we use the term *basic knowledge theories* to refer to epistemological theories that do not reject the possibility of *basic knowledge*. Basic knowledge, in this context, is the knowledge that is obtained from a reliable source without the agent having the prior knowledge that the source is indeed reliable. The term *basic knowledge theories* is meant to be an umbrella term that applies to many approaches to knowledge, most of which are considered externalist or naturalist in one way or another. Examples for such theories include reliabilism, safety theories, sensitivity theories, anti-luck theories, and causal theories of knowledge.[14] Common to these theories is the idea that the agent does not need to know that a source of information is reliable in order to actually gain knowledge from that source.

Let us complicate our simple example from Figure 2 to the following toy example.

---

[14] For a survey of these theories, see Ichikawa and Steup (2018). See also Lyons (2016) for a related discussion about *modest foundationalism*.

> **Clock Tower:** Ann is visiting a foreign village. She looks at a clock tower that points to 12:00. The clock tower has, in fact, a perfectly reliable clock mechanism, although Ann does not know that (neither before looking at the clock nor after).

Basic knowledge theories will gladly accept that it is possible for Ann to come to know that the time is 12:00 as a result of looking at the clock (that would be basic knowledge).

Theories that reject basic knowledge argue that since Ann does not know that the clock mechanism is reliable, she does not come to know the time. Both sides of this debate agree that if that clock mechanism is in fact unreliable, then Ann does not come to know the time by looking at the clock.

We can take a dynamic perspective on this debate. Both sides agree on an initial epistemic situation in which Ann does not know that the time is 12:00 (denote that proposition p) and does not know that the clock mechanism is reliable (denote that with r). The basic knowledge theorist describes the event of looking at the clock as the event s.t. if the clock mechanism is reliable (r) then Ann comes to know p, while if the clock is unreliable then she comes to know nothing new. The theorist that rejects basic knowledge describes the event of looking at the clock as the event in which no new knowledge about the time is gained, no matter what (because, as both sides assumed, Ann does not know that the clock mechanism is reliable).

It is worth highlighting here the way many externalists have come to understand the notion of *evidence*, and the way it functions in examples like the one above. Contrast two scenarios: in the first the clock is broken (and hence unreliable); in the second the clock is reliable. Intuitively, when Ann looks at the clock that points to 12:00, she receives the same evidence in both scenarios. Externalists tend to reject this intuition. According to many contemporary externalists, evidence is factive: evidence is always true and false evidence is not real evidence.[15] If evidence is factive, then Ann does not receive the same evidence in both scenarios: if the clock is in fact broken, Ann does not have evidence that the time is 12:00; if the clock is reliable, she does. Externalists defend the claim that evidence is factive by offering an 'external' analysis of evidence, under which the status of evidence is affected by factors that go beyond the agent's intrinsic mental state (like whether the source of the information is reliable or not). Externalists can accept that in both scenarios Ann comes to believe that the time is 12:00, even if she has different evidence in the two scenarios. This picture, of course, complicates the relation between one's

---

[15] See Williamson (2000), Littlejohn (2013), Bird (2018), Neta (2018), Fratantonio and McGlynn (2018), and Salow (2017) for a factive take on evidence. See Goldman (2009), Rizzieri (2011) and Comesaña and Kantin (2010) for objections. My models do not strictly speaking assume that evidence is factive, because the models do not explicitly model evidence. But I do think that the models help frame this debate.

belief, evidence and knowledge, and remains a topic of debate within epistemology.[16] The agent's doxastic state will not be modeled in this section.

In the clock tower example, we assume that Ann does not know if the clock is reliable or not. Hence, according to the externalist, she cannot predict what she will learn (or what evidence she will receive) as a result of looking at the clock. I claim that we can model the externalist understanding of the event of Ann looking at the clock as the program $(?r; !p) \cup (?\neg r; !\top)$. We read this program as the following event: if r is the case (the clock mechanism is reliable), transition into a situation in which Ann knows p (the time is 12:00); if r is not the case, transition into a situation in which Ann has not learned anything new. Figure 5 depicts this type of update with a forest model.

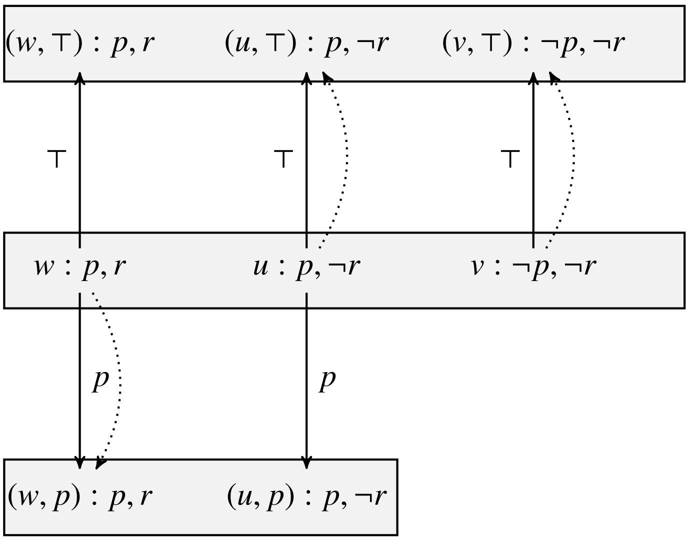

Figure 5: A basic knowledge conception of learning p without knowing that the source is reliable — the dotted relation.

The middle box in Figure 5 represents Ann's initial situation: she does not know p (that the time is 12:00) nor does she know r (that the clock mechanism is reliable). The dotted relation depicts the program (or event) $(?r; !p) \cup (?\neg r; !\top)$, which is the event of learning p if r is the case and learning nothing if not r is the case. Observe the depiction of that event, as the dotted relation, in Figure 5: in the world in which r is true (world w, the clock mechanism is indeed reliable), the dotted relation sends us to a submodel in which p is known (at the bottom). In the worlds in which the clock mechanism is unreliable (worlds u and v) the dotted relation sends us to a submodel in which nothing has changed (the top model). The event of looking at the clock has different epistemic consequences, depending on which world is actual.

---

[16] This problem has been labeled the *new evil demon problem.* For an overview, see Littlejohn (2009) and Beddor and Goldman (2015).

A few observations about the dotted relation, the epistemic event $(?r; !p) \cup (?\neg r; !\top)$: first, note that it does not fit the framework depicted in Figure 1, it is an update which is sensitive to the world of evaluation (the external environment); it is not just a function from one model to the other. Second, note that this event is deterministic: at each particular environment it has a deterministic result.[17] Graphically, the relation is a partial function (a function in fact). Third, note that the dotted relation violates the semantic condition which corresponds to the NM axiom: from w we can get to (w, p) via the dotted relation, from w we can get to u via the R relation (which, recall, is depicted as the grey boxes), and from u we can get to $(u, \top)$ via the dotted relation. But we cannot get from world (w, p) to world $(u, \top)$ via the R relation, as the semantic NM condition would require.

Thus, the event $(?r; !p) \cup (?\neg r; !\top)$ depicted in Figure 5 offers a counter example to the NM axiom, stating that $\langle\pi\rangle K\varphi \rightarrow K[\pi]\varphi$. On the one hand, it is the case that $M, w \models \langle(?r; !p) \cup (?\neg r; !\top)\rangle Kp$: in world w, as a result of looking at the clock, Ann comes to know that the time is 12:00 (according to the basic knowledge theorist). On the other hand it is also the case that $M, w \models \neg K[(?r; !p) \cup (?\neg r; !\top)]p$: prior to looking at the clock, Ann does not know that looking at a clock pointing to 12:00 implies that the time is in fact 12:00, since, recall, Ann does not know that the clock mechanism is reliable. For all Ann knows, the actual world is v, in which the time is not 12:00 and the clock mechanism is unreliable. In such a world, looking at a clock that points to 12:00 does not imply that the time is 12:00. Thus, the basic knowledge theorist that understand the event of looking at the clock as the dotted relation rejects the NM principle.[18]

The event $(?r; !p) \cup (?\neg r; !\top)$ is opaque: prior to its execution, Ann does not know what its result will be. Ann does not know if she is in the good case or the bad case, so the effect of looking at the clock is not transparent to her. The event fails the NM principle because Ann does not know where she is going, epistemically speaking.

The objector to basic knowledge uses the contra-positive form of NM to reject the possibility of an event like $(?r; !p) \cup (?\neg r; !\top)$. According to the objector, since Ann does not know that the event π results in a p state (you don't know that looking at a clock that says the time is 12:00 implies that the time is actually 12:00), then the event π does not result in knowledge that p (you don't really know that the time is 12:00 after looking at the

---

[17] Thus, the opaqueness of that event can only be attributed to the agent's ignorance, not to any ontic chances.

[18] As I mentioned in the introduction, externalism is often assumed to be inconsistent with the KK principle. See Okasha (2013), Bird and Pettigrew (2019) for recent discussions and Williamson (2000) for the locus classicus. The picture painted here complicates the discussion. I have argued that it is natural to understand externalists as rejecting NM first and foremost. Since KK and NM are logically independent, there is nothing inconsistent about rejecting the one while endorsing the other. I believe that it is worthwhile re-evaluating familiar arguments against introspective knowledge given the distinction between static and dynamic introspection I draw here. This issue will not be developed in this paper.

clock). This is the implication $\neg K[\pi]p \rightarrow \neg\langle\pi\rangle Kp$, which is the contra-positive form of NM, $\langle\pi\rangle Kp \rightarrow K[\pi]p$. We can understand the objector to basic knowledge as endorsing NM, and with it, a transparent conception of updates (if you don't know that the source is reliable, you can't get knowledge out of it).

Upshot: since the rejection of NM allows agents to gain knowledge from sources they don't know to be knowledge conducive, the truth of NM plays a key role in debates about basic knowledge.[19] PAL with opaque updates allows us to model both sides of this debate.

3.2 The bootstrapping problem

The bootstrapping problem is a famous objection that is usually raised against basic knowledge theories.[20] According to objectors, basic knowledge theorists allow agents to learn about the reliability of their sources of information for free, or too easily, after the update, without sufficient evidence for doing so. This way of learning amounts to illicit bootstrapping, or easy knowledge, according to the objector. The bootstrapping problem comes in many variations and has many interesting proposed solutions.[21] The discussion here is not meant to offer a decisive solution to the problem, but to show that opaque updates, and in particular the PR principle, are quite relevant to at least one formulation of the problem.

We can present a toy version of the problem with our clock example modeled in Figure 5.[22] The objector to basic knowledge theories notes that initially (in the middle submodel of Figure 5) Ann knows that as a result of looking at the clock, if she comes to know the time, then the clock mechanism must be reliable. In other symbols, the formula $K[\pi](Kp \rightarrow r)$ is true at world w, where $\pi = (?r; !p) \cup (?\neg r; !\top)$, the event of looking at the clock according to the basic knowledge theorist. There is no world in the model in Figure 5 accessible via the R_π relation in which $Kp \wedge \neg r$ is true, so we get $M, w \models K[\pi](Kp \rightarrow r)$. Now, the objector argues informally as follows: we agree that Ann has the prior knowledge that as a result of looking at the clock, if she knows the time, then the clock mechanism must be reliable. Further, according to the basic knowledge theorists, Ann can come to know the time after actually looking at the clock. Thus, what stops Ann from putting the two pieces of information together and concluding, after

---

[19] The philosophical connections between No-Miracles, basic knowledge, and skepticism are further developed in Cohen (2020).

[20] See Vogel's attack on reliabilism (2000, 2008) and Cohen (2002). Some take the problem to be more widespread (Weisberg 2010, van Cleve 2003). Analogous problems exist for justification and belief (White 2006, Pryor 2013, Weatherson 2007).

[21] See, e.g., the survey in Weisberg (2012).

[22] This example of bootstrapping does not involve any inductive elements, unlike Vogel's original version of the problem (2000, 2008). However, as Titelbaum (2010) has argued, induction is not essential to the bootstrapping problem.

looking at the clock, that the clock mechanism is in fact reliable (illicitly bootstrapping her way to the knowledge that the clock is reliable, so to say)? The problem is that in order to learn that the clock mechanism is reliable, it seems, one needs to do more than merely glance at it. Since the conclusion that Ann comes to know that the clock is reliable just by looking at the clock face is absurd, something must be wrong with the informal argument. The objector suggests rejecting the basic knowledge theorist's assumption that Ann learns the time by looking at the clock.

Here is one way to make the above informal argument more formal: we assume $[\pi]Kp$ (Ann knows the time as a result of looking at the clock) and $K[\pi](Kp \to r)$ (as in the last paragraph). Using the PR principle, $K[\pi]\varphi \to [\pi]K\varphi$, we can conclude $[\pi]K(Kp \to r)$, stating that as a result of looking at the clock, Ann knows that if she knows the time, the clock mechanism must be reliable. Given the KK principle and the closure principle of knowledge, from $[\pi]Kp$ and $[\pi]K(Kp \to r)$ we can deduce $[\pi]Kr$: as a result of looking at the clock, Ann knows that the clock is reliable. To summarize:

15. $[\pi]Kp$ basic knowledge assumption
16. $K[\pi](Kp \to r)$ assumption about Ann's background knowledge
17. $[\pi]K(Kp \to r)$ by PR from 16.
18. $[\pi]KKp$ instance of KK on 15.
19. $[\pi]Kr$ by the closure of knowledge from 17. and 18.

So, my formulation of the bootstrapping argument requires assuming an instance of the closure of knowledge, the KK principle, and the PR principle.[23] Note that the model in Figure 5 makes sentences 15. and 16. true (at world w) and assumes both the KK principle (the transitivity of the R relation) and the closure principle of knowledge, but $[\pi]Kr$ is false: after looking at the clock (world (w, p)) Ann does not know r. The bootstrapping argument is blocked in Figure 5 because the update π, apart from violating NM, further violates PR. From world w one can go first to world (w, p) via the π relation (dotted), then to world (u, p) via the R relation, but there is no way to get from world w to world (u, p) first via the R relation and then via the dotted π relation. Syntactically, note that $K[\pi](Kp \to r) \land \neg[\pi]K(Kp \to r)$ (which is true in M, w) is a counterexample to the PR principle.

The model in Figure 5 violates the PR principle not because Ann forgets anything after looking at the clock. The failure of PR occurs because Ann, even after the epistemic event

---

[23] The formulation also assumes that $[\pi]$ is a normal modal operator. Recall that we indeed treat π as a normal modal operator in the logic of opaque updates. I don't see how this assumption would be challenged on epistemological grounds.

occurred, does not know what event has brought her to her current epistemic situation. Consult Figure 5 again, and assume that w is the actual world. We can see that both $K[\pi](Kp \rightarrow r)$ and $[\pi]Kp$ are true. However, after the event of looking at the clock (world (w, p)), Ann considers it possible that the actual world is (u, p), meaning she considers it possible that the initial actual world is u and that she learned that p at that world. This type of ignorance can be cashed out as a difference between *de-re* and *de-dicto* knowledge. Ann can have the de-dicto knowledge that as a result of the event *looking at a clock tower*, if she knows the time, then the clock mechanism must be reliable. At the same time, she can be ignorant of the fact that *that* event is the event that resulted in her current knowledge (this is *de-re* ignorance). The ignorance expressed in the failure of PR arises from the possibility that different distinct reliable sources could have given the agent the knowledge they have. For a concrete, but somewhat outlandish example, take world u to be a world in which the clock mechanism is unreliable but in which a benevolent God transplants the correct time in Ann's mind. Since Ann cannot rule out this non-actual epistemic event (maybe she just read Descartes' *Meditations*), she cannot rule out the possibility that an epistemic event different than the actual one produced her knowledge. The epistemic event that Ann experienced is opaque in the sense that she cannot identify it as the event from which she obtained knowledge about the time. This is why world (u, p) cannot be ruled out after the update.[24]

Existing discussions about the bootstrapping problem seem to be unaware of the fact that PR plays an essential role in bootstrapping reasoning. We see that externalists can dissolve the bootstrapping problem by noting that an update like the clock tower update is opaque, and therefore incompatible with PR. In general, PR will fail in situations in which agents don't know how they know φ (or what is the epistemic event that resulted in knowing φ), even if they know φ. Again, basic knowledge theories are more open to the possibility of such situations, so such theories could be understood as rejecting both PR and NM. The simple PAL forest model of Figure 5 can represent this complicated epistemic scenario.

4 Comparison to other frameworks and further work

Opaque updates, as an epistemic-logical phenomenon, are not exclusive to the particular logical system I have presented here. Sections 2 and 3 presented and used a simple PDL-PAL hybrid system that has the advantage of being relatively close to the (now very) familiar PAL, while flexible enough to easily model the failure of both NM and PR. In this

---

[24] Further note that if world (u, p) would be eliminated from the model in Figure 5, then PR would be true in w, while NM would still be false. One way of formally interpreting the bootstrapping objection is as the insight that rejecting NM while endorsing PR is a strange epistemological combination.

Section, I discuss other logical frameworks that highlight interesting features of opaque updates:

1. Multi-pointed event models: One standard way to generalize PAL is to model updates via *event models* that can represent a wide range of communication events. In DEL with event models, an updated epistemic model is the product $M \times A$ of an initial epistemic model M and an event model A. An event model is a Kripke model with a finite set of events E, an accessibility relation between events, and a precondition function for each event $e \in E$, which intuitively specifies which formula is announced for each e. An event modality $[A, e]\varphi$ is added to the language, stating that φ holds as a result of executing event e. See Baltag and Renne (2016) for a proper overview on event models.

Consider the event model A with two events, e_1 and e_2, such that the precondition of e_1 is p and the precondition of e_2 is $\top$. In the context of the clock tower example, we can think of $[A, e_1]$ as the event of looking at a reliable clock (resulting in knowledge of p) and of $[A, e_2]$ as the event of looking at an unreliable clock (resulting in no new knowledge). $[A, e_1]$ and $[A, e_2]$ are not opaque updates, intuitively because they specify which event is actual (e_1 or e_2). We can, however, easily build opaque updates out of them by using the *multi-pointed* event operator $[A, E]$ (see Sietsma and van Eijck (2012), Baltag and Renne (2016)).[25] The *multi-pointed* event operator $[A, E]\varphi$ abbreviates the conjunction $\bigwedge_{e \in E} [A, e]\varphi$. Likewise, the diamond operator $\langle A, E\rangle$ stands for the disjunction $\bigvee_{e \in E} \langle A, e\rangle\varphi$.

The operator $\langle A, E\rangle$ is non-deterministic and opaque, since the agent cannot know which event e_i it executes. Taking the middle model of Figure 5 as our initial epistemic model, we have the following No-Miracles failure: $M, w \models \langle A, E\rangle Kp$ and $M, w \models \neg K[A, E]p$. Note, however, that there is no direct equivalence between this approach to opaque updates and the one I presented in Sections 2-3. Recall that conceptually, there is a difference between whether an update is deterministic and whether the agent can predict its behavior. The latter is an epistemic issue; the former is ontic. The program π from Section 3 is deterministic, even though the agent cannot predict its behavior. The update $\langle A, E\rangle$ on the other hand is truly non-deterministic (note that we have $M, w \models \langle A, E\rangle \neg Kp \wedge \langle A, E\rangle Kp$), which explains why the agent cannot predict its behavior. This difference also implies that we cannot directly interpret the formula $K[A, E]\varphi \rightarrow [A, E]K\varphi$ as capturing the epistemic intuition behind Perfect-Recall (in the clock example, the sentence $K[A, E](Kp \rightarrow r) \rightarrow [A, E]K(Kp \rightarrow r)$ is true as its antecedent is false). A richer study of opaqueness using the framework of event models is of course welcomed, and I leave it for future investigation.

2. The dynamics of knowing a value: There is a growing literature on dynamic-

---

[25] I thank an anonymous reviewer for suggesting this connection.

epistemic logics for wh-knowledge, like knowing *what* is the value of a variable.[26] Consider the event of announcing the truth value of p. This event is opaque in that it violates No-Miracles: if p is true and the agent does not know it, the agent cannot predict that they will know p after the event of announcing the truth value of p. The agent can predict, however, that after the event they will know the value of p, so they are not ignorant about the fact the event is successful in conveying the value of p (this is unlike the examples from Section 3). Epistemic logics that go beyond knowing-that can provide further stimulating perspectives on opaqueness.

3. Plausibility models for doxastic epistemic logics: Existing dynamic doxastic epistemic logics offer an impressive model for formalizing an externalist notion of knowledge as belief that is stable under revision with true information. The framework of Baltag and Smets (2008) in particular, succeeds in combining different notions of belief, knowledge, update, and belief revision in a unified manner. The *radical upgrade* operator $\Uparrow \varphi$ of that system represents the agent's belief revision with the (possibly false) φ, and is modeled with a doxastic plausibility ordering. The system includes two notions of knowledge: an S4 type modality K representing defeasible knowledge, and a stronger S5 modality $\Box$ representing irrecoverable knowledge. Defeasible knowledge is equivalent to belief that is stable under revision with true information.

It is quite interesting to study opaqueness relative to different modalities and updates in this system. For example, the NM principle $\langle \Uparrow \varphi \rangle K\psi \to K[\Uparrow \varphi]\psi$ is not valid in the system: an agent will always come to know p after a radical upgrade with p, assuming that p is true, i.e. $p \to \langle \Uparrow p \rangle Kp$. But if the agent initially does not believe that p, then they will not know that revising p implies that p is actually the case: $\neg Bp \to \neg K[\Uparrow p]p$. This failure of NM nicely captures the externalist elements of the defeasible knowledge operator K: the epistemic result of an update depends on factors the agent is initially ignorant about.

Plausibility models also provide an excellent tool to further study the effects of false evidence from an externalist perspective. I have used the program $\pi = (?r; !p) \cup (?\neg r; !\top)$ to model the clock tower example. This modeling is incomplete, since it ignores the (quite plausible) assumption that regardless of the reliability of the clock (the truth of r), the agent comes to believe the time is 12:00 (i.e. p) as a result of looking at the clock (even if they don't know p). We can model the clock tower example with plausibility models, and consider a program like $\pi' = ((?r; !p) \cup (?\neg r; !\top); \Uparrow p)$. This program will have different effects on the agent's irrecoverable knowledge, but it results in belief in p come what may.[27]

---

[26] See Wang (2018) for a broad overview, and van-Eijck et al. (2017), Baltag (2016) for dynamic epistemic logics of knowing a value.

[27] As I briefly mentioned in the introduction, I chose to model opaque updates with the simple PDL-

4. Topological models of PAL: The No-Miracles and Perfect-Recall principles correspond to interesting topological properties in topological semantics of epistemic logic. In particular, if we think of the $\rightarrow_\pi$ relation in forest models as a partial function between points of a topological structure, then the formula $\langle\pi\rangle K\varphi \rightarrow K[\pi]\varphi$ corresponds to the continuity property of π. See Kremer and Mints (2005), and Bjorndahl (2018) for works relevant to the connection between continuity and transparent updates.

5 Concluding remarks

Taking a dynamic perspective, we can think of epistemic internalism as the claim that posterior epistemic states *supervene on the agent's prior epistemic state*. Indeed, in some formal frameworks (like Bayesian update), sentences about the posterior epistemic state can be *reduced* to sentences about the agent's prior epistemic state. Externalists, on the other hand, hold that the effects of epistemic events supervene both on the epistemic state of the agent *and* on environmental conditions external to the agent. Opaque updates allow us to model the externalist idea that the result of an epistemic event may depend on non-epistemic features of the environment. An opaque update U can result in knowledge that p in a world where the non-epistemic fact r (which is logically independent of p) is true and can also result in no new substantial knowledge in a world in which r is false. As we have seen, the semantics of such opaque updates directly correspond to the failure of the (syntactic) NM and PR principles. It is thus valuable to understand externalist theories as rejecting NM and PR.

From our current perspective, it is fair to say that Hintikka's (1962) seminal work in epistemic logic has *reshaped* the way many epistemologists think about introspection. Static positive and negative introspection are now taken for granted as tools in the epistemologist's toolbox, whether people accept them or not. My hope is that, likewise, recent developments in dynamic epistemic logic will offer epistemologists the conceptual tools to think about *dynamic* forms of introspection in a precise and simple manner.

Acknowledgments I would like to thank Johan van Benthem, Ray Briggs, and Krista Lawlor for their tremendous help in the process of developing and writing this paper. I also thank two anonymous reviewers of this journal for many helpful comments, suggestions, and corrections on an earlier version of this paper. Their valuable comments helped in clarifying the goal and scope of this paper.

PAL hybrid system of Section 2 because (unlike plausibility models) it remains neutral with respect to the truth of the KK principle and the exact relations between knowledge and belief. Both of these issues are controversial in contemporary philosophical discussions of externalism.

References

– Arntzenius, Frank (2003). Some Problems for Conditionalization and Reflection. *Journal of Philosophy*. 100 (7):356-370.

– Aucher G., Herzig A. (2010). Exploring the power of converse events. in *Dynamic Epistemology: Contemporary Perspectives*, Springer, Synthese Library.

– Baltag, Alexandru (2016). To Know is to Know the Value of Variable. In Lev Beklemishev, Stéphane Demri & András Máté (eds.), *Advances in Modal Logic*, Volume 11. CSLI Publications. pp. 135-155.

– Baltag, Alexandru and Renne, Bryan. (2016) Dynamic Epistemic Logic. *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition).

– Baltag, Alexandru and Smets, Sonja. (2008) A qualitative theory of dynamic interactive belief revision, in G. Bonanno, W. van der Hoek, and M. Wooldridge (eds.), *TLG 3: Logic and the Foundations of Game and Decision Theory (LOFT 7)*, Volume 3 of Texts in logic and games, pp. 11–58, Amsterdam: Amsterdam University Press.

– van Benthem, Johan; Gerbrandy, Jelle; Hoshi, Tomohiro & Pacuit, Eric (2009). Merging frameworks for interaction. *Journal of Philosophical Logic* 38 (5):491-526.

– van Benthem, Johan (2012). *Logical Dynamics of Information and Interaction.* Cambridge University Press.

– van Benthem, Johan (2014). *Logic in Games*. Cambridge, MA: The MIT Press.

– van Benthem, Johan and Klein, Dominik. Logics for Analyzing Games., *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition), Edward N. Zalta (ed.).

– van Benthem, J., J. van Eijck, and B. Kooi. (2006) Logics of communication and change, *Information and Computation,* 204(11): 1620–1662.

– Bird, A. (2018), Evidence and Inference. *Philosophy and Phenomenological Research*, 96: 299-317.

– Bird, Alexander & Pettigrew, Richard (2019). Internalism, Externalism, and the KK Principle. *Erkenntnis*:1-20.

– Bjorndahl A. (2018) The Epistemology of Nondeterminism. In: Moss L., de Queiroz R., Martinez M. (eds) *Logic, Language, Information, and Computation.* WoLLIC 2018. Lecture Notes in Computer Science, vol 10944. Springer, Berlin, Heidelberg

– Carter, A.J., Gordon. C.E and Jarvis B. (eds), (2017). *Knowledge First: Approaches in Epistemology and Mind.* Oxford University Press.

– van Cleve, James. (2003) Is knowledge easy — or impossible? externalism as the only alternative to skepticism. In *The Skeptics: Contemporary Essays. Aldershot, UK: Ashgate Publishing.*

– Cohen, Michael (2020). The problem of perception and the no-miracles principle. *Synthese*. https://doi.org/10.1007/s11229-020-02772-3

– Cohen, Stewart (2003). Basic Knowledge and the Problem of Easy Knowledge. *Philosophy and Phenomenological Research*, 65(2): 309–329.

– Comesaña, J., and Kantin, H. (2010). Is evidence knowledge? *Philosophy and Phenomenologi-*

cal Research, 80, 447–454.

– van Ditmarsch, H.P. *Knowledge games.* Ph.D. thesis, University of Groningen, 2000. ILLC Dissertation Series DS-2000-06.

– van Eijck J., Gattinger M., Wang Y. (2017) Knowing Values and Public Inspection. In: Ghosh S., Prasad S. (eds) *Logic and Its Applications. ICLA* 2017. Lecture Notes in Computer Science, vol 10119. Springer, Berlin, Heidelberg.

– Fratantonio, G., and McGlynn, A. (2018). Reassessing the Case against Evidential Externalism. In V. Mitova (Ed.), *The Factive Turn in Epistemology* (pp. 84-101). Cambridge: Cambridge University Press.

– Girard, P., F. Liu, and J. Seligman (2012). General dynamic dynamic logic. In T. Bolander, T. Brauner, S. Ghilardi, and L. Moss (Eds.), *Proceedings of the 9th International Conference on Advances in Modal Logic (AiML'12)*, pp. 239–260. London: College Publications.

– Goldman, A. (2009). Williamson on knowledge and evidence. In P. Greenough and D. Pritchard (ed.), *Williamson on knowledge* (pp. 73-92). Oxford University Press.

– Halpern, Joseph (2004). Sleeping Beauty Reconsidered: Conditioning and Reflection in Asynchronous Systems. In Tamar Szabo Gendler & John Hawthorne (eds.), *Proceedings of the Twentieth Conference on Uncertainty in Ai.* Oxford University Press. pp. 111-142.

– Hintikka, Jaakko (1962). Knowledge and Belief. New York: Cornell University Press.

– Kremer, Philip and Mints, Grigori. (2005). Dynamic topological logic. *Annals of Pure and Applied Logic* 131 133–158.

– Lyons, Jack, (2016) Epistemological Problems of Perception, *The Stanford Encyclopedia of Philosophy*

– Ichikawa, Jonathan Jenkins and Steup, Matthias. (2018) The Analysis of Knowledge. *The Stanford Encyclopedia of Philosophy*(Summer 2018 Edition), Edward N. Zalta (ed.).

– Littlejohn, Clayton (2009). The New Evil Demon Problem. *Internet Encyclopedia of Philosophy.*

– Littlejohn, Clayton (2013). No Evidence is False. *Acta Analytica* 28 (2):145-159.

– Miller, J.S. and L.S. Moss. (2005). The undecidability of iterated modal relativization. *Studia Logica*, 79(3): 373–407.

– Nagel, Jennifer (2013). Knowledge as a Mental State. *Oxford Studies in Epistemology*. 4:275-310.

– Neta, R. (2018). Your Evidence Is the Set of Facts That Are Manifest to You. In V. Mitova (Ed.), *The Factive Turn in Epistemology* (pp. 32-49). Cambridge: Cambridge University Press.

– Okasha, S. (2013). On a flawed argument against the KK principle. *Analysis*, 73(1), 80–86.

– Pettigrew, Richard (2019). What is conditionalization, and why should we do it? *Philosophical Studies*:1-37.

– Pryor, James (2013). Problems for Credulism. In Chris Tucker (ed.), *Seemings and Justification: New Essays on Dogmatism and Phenomenal Conservatism.*

– Rizzieri, A. (2011). Evidence does not equal knowledge. *Philosophical Studies*, 153, 235–242.

– Rott, H. (2004) Stability, Strength and Sensitivity: Converting Belief into Knowledge. *Erkennt-*

nis. 61, 469–493.
– Salow, Bernhard (2018). The Externalist's Guide to Fishing for Compliments. *Mind.* 127 (507):691-728.
– Sietsma, F. and J. van Eijck. (2012). Action emulation between canonical models, in *Proceedings of the 10th Conference on Logic and the Foundations of Game and Decision Theory (LOFT 10)*, Sevilla, Spain.
– Stalnaker, R. (2006). On Logics of Knowledge and Belief. *Philosophical Studies* 128, 169–199.
– Titelbaum, M.G. (2010). Tell me you love me: bootstrapping, externalism, and no-lose epistemology. *Philosophical Studies*. 149, 119–134
– Troquard, Nicolas and Balbiani, Philippe (2019). Propositional Dynamic Logic. *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition), Edward N. Zalta (ed.)
– Vogel, Jonathan (2000). Reliabilism Leveled. *The Journal of Philosophy*. 97(1): 602–623.
– Vogel, Jonathan (2008). Epistemic Bootstrapping. *The Journal of Philosophy*. 105 (9):518-539.
– Wang, Y., Kuppusamy, L., & van Eijck, J. (2009). Verifying epistemic protocols under common knowledge. In TARK '09 (pp. 257–266).
– Wang, Yanjing & Cao, Qinxiang (2013). On axiomatizations of public announcement logic. *Synthese* 190 (S1).
– Wang, Yanjing (2018). Beyond Knowing That: A New Generation of Epistemic Logics. In Hans van Ditmarsch & Gabriel Sandu (eds.), *Jaakko Hintikka on Knowledge and Game Theoretical Semantics*. Springer. pp. 499-533.
– Weatherson, Brian (2007). The Bayesian and the Dogmatist. *Proceedings of the Aristotelian Society* 107:169-185.
– Weisberg, Jonathan (2010). Bootstrapping in General. *Philosophy and Phenomenological Research*. 81 (3):525-548.
– Weisberg, Jonathan (2012). The Bootstrapping Problem. *Philosophy Compass* 7 (9):597-610
– White, Roger (2006). Problems for Dogmatism. *Philosophical Studies* 131 (3):525-557.
– Williamson, Timothy (2000). *Knowledge and its Limits*. Oxford University Press.
– Williamson, Timothy (2013). Gettier cases in epistemic logic. *Inquiry*, 56, 1–14.
– Williamson, Timothy (2014). Very improbable knowing, *Erkenntnis*. 79, 971–999.
– Yap, Audrey, Hoshi, Tomohiro. (2009) Dynamic Epistemic Logic and Branching Temporal Structures. *Synthese* 169 (2009): 259-281.