# Inexact Knowledge and Dynamic Introspection

**Abstract**

Cases of inexact observations have been used extensively in the recent literature on higher-order evidence and higher-order knowledge. I argue that the received understanding of inexact observations is mistaken. Although it is convenient to assume that such cases can be modeled statically, they should be analyzed as dynamic cases that involve change of knowledge. Consequently, the underlying logic should be dynamic epistemic logic, not its static counterpart. When reasoning about inexact knowledge, it is easy to confuse the initial situation, the observation process, and the result of the observation; I analyze the three separately. This dynamic approach has far reaching implications: Williamson's influential argument against the KK principle loses its force, and new insights can be gained regarding synchronic and diachronic introspection principles.

## 1 Introduction

According to *externalist* theories of knowledge, the factors that make knowledge different from mere true belief might be external, and so inaccessible, to the epistemic subject. Externalist theories thus seem to be in tension with the introspective capacities of epistemic subjects. This tension plays a key role in Timothy Williamson's work on perceptual knowledge and the failure of the KK principle, according to which *S knows that p* entails *S knows that S knows that p* (Williamson 2000). The cases that motivate Williamson are cases of *inaccurate* or *inexact knowledge*, which emerge whenever we gain knowledge from our imperfect, often inaccurate, perceptual capacities.

Cases in which the KK principle fails, according to the Williamsonian picture that emerges, can give rise to more extreme cases in which our second-order epistemic attitudes radically differ from our first-order attitudes: it is possible to know $p$ while being extremely confident, given one's evidence, that one does not know $p$ (Williamson 2014). Given the right (better yet, wrong) evidential state, our first-order epistemic life might be completely foreign to us. This is Williamson's story.

This story has influenced the way epistemologists have recently approached the question of higher-order evidence: in what ways should our higher-order evidence relate to our lower-order evidence? (Christensen, 2010). According to *modesty*, it is sometimes rational not to be fully confident with regards to the question "what should my level of confidence be?" Rejecting modesty does not seem like a privilege that fallible creatures like us enjoy. But finding the correct way to combine first- and second-order evidential attitudes has proven to be a non-trivial philosophical task. There seems to be, however, agreement in the literature that cases of inexact knowledge, and more generally inexact observations, are cases in which conflicts between first- and higher-order evidence emerge.[1]

The general Williamsonian story has also been used to argue against the possibility of common knowledge (Lederman 2017), Good's Theorem in Bayesian epistemology (Salow & Ahmed, 2017), the Stalnakerian picture of assertion at the foundation of formal semantics (Hawthorne & Magidor, 2009), and standard assumptions in the epistemology of indicative conditionals (Holguín 2019), to name just a few applications. The Williamsonian account of inexact knowledge has proven to be extremely influential. At the same time, the KK principle — the rejection of which lies at the heart of the Williamsonian story — keeps having its defenders.[2] The debate is far from dead.

Here I want to tell a different story about the tension between introspection and inexact observations. I argue that Williamson's understanding of scenarios of inexact perceptual knowledge is incomplete as it stands and, in particular, fails to show that the KK principle is false.[3] On the contrary, inexact perceptual observations are compatible with the KK principle, once the logical mechanism with which agents *update* their knowledge after they make inexact observations is clarified. Situations of inexact learning are not static, they are situations in which change of knowledge occurs due to a perceptual event. Consequently, I argue, the underlying logic of inexact observations should be *dynamic epistemic logic*, not its static counterpart. Cases of inexact observations do *not* force a (synchronic) conflict between first-

[1] See Horowitz (2014), Elga (2013), Dorst (2020), Lasonen-Aarnio (2014, 2015), Roush (2017).

[2] For recent defense of the KK thesis, see Das and Salow (2016), Dorst (2019), Goodman and Salow (2018), Greco (2014a,2014b, 2015a, 2017), Stalnaker (2015).

[3] Williamson's argument is arguably the most influential criticism of the KK principle in contemporary epistemology, but it is not the only one. My analysis does not address other arguments against the KK principle, such as Liu (2020) and the arguments reviewed in Greco (2015b).

and higher-order knowledge. At the same time, I argue, when inexact observations occur, epistemic agents cannot know how *future* evidence will affect their knowledge state, even if they are fully (synchronically) introspective. Inexact observations show that epistemic agents do not always have dynamic (or diachronic) introspection.[4] This kind of diachronic uncertainty is worth further inquiry, especially in the context of externalist theories of knowledge and evidence. I argue that the real conflict is between inexact observations and dynamic introspection, not the KK principle.

To my knowledge, nearly all earlier critiques of Williamson's argument against the KK principle have ignored the dynamic nature of inexact observations.[5] It is thus valuable to examine that dynamic aspect thoroughly, which is my aim here. In the rest of this section, I present Williamson's own understanding of inexact knowledge. I explain Williamson's understanding and formulation of the *margin-for-error principle*, which is central to his way of thinking about inexact knowledge. In Section 2, I develop my dynamic account of inexact knowledge and present a natural way to syntactically enrich Williamson's original argument in a dynamic language. Given my alternative dynamic reconstruction, a tension arises between a *dynamic introspection principle* and my dynamic formulation of the margin-for-error principle. This offers an alternative explanation to the tension between inexact observations and introspection. Section 3 contains my novel semantics for inexact updates, which enriches the syntactic analysis of Section 2 and sketches a general account of the epistemology of inexact observations. Section 4 reevaluates Williamson's margin-for-error principle. I argue that given my alternative dynamic explanation, Williamson's *static* formulation of the margin-for-error principle should be rejected. The margin-for-error principle is a principle about knowledge *obtained from an inexact observation*, not about knowledge *in general*; Williamson's account fails to make this distinction. My dynamic account is able to capture the nature of inexact observations, including the motivation behind the

[4]This is true for both logical and probabilistic formulations of introspection. One can add *evidential probabilities* in the style of Williamson (2014) to the epistemic models I present here. It can then be shown that on such models the probabilistic *diachronic reflection principle* fails when inexact observations occur. This is analogous to Williamson's (2014) demonstration that the probabilistic (synchronic) *reflection principle* fails on Williamson's static models. I leave this out due to space limitations.

[5]For these non-dynamic critiques, see Mott (1998), Brueckner and Fiocco (2002), Neta and Rohrbaugh (2004), Conee (2005), Dutant (2007), Greco (2014a) Halpern (2008), Egre and Bonnay (2008, 2009, 2011), Sharon and Spectre (2008) and Stalnaker (2015). Egre and Bonnay use dynamic epistemic logic, but not in order to model the act of observation. The only exception is Baltag and van-Benthem (2018), who take a dynamic approach different than mine.

margin-for-error principle, while avoiding the problems that Williamson's original picture faces. Section 5 concludes.

### 1.1 Williamson's Argument

For the sake of familiarity, I use the unmarked clock example (Williamson 2014, Elga 2013) as my guiding example in this paper. Since I will end up offering a general account of inexact observations, my analysis can be generalized to other similar examples. Here is the scenario: we have an analog clock that lacks marks for hours and minutes. The hands of the clock point to 12:17. Ann is looking at the clock from afar. Since Ann has normal human perceptual abilities, it is not the case that after looking at the clock Ann knows that the clock is pointing at 12:17. However, Ann does learn something from looking at the clock. Ann has a margin-for-error, s.t. if the minute hand is pointing to minute $i$, then for all Ann knows it points to $i \pm 1$. That is, Ann's margin-for-error is 1 minute. Thus, after looking at the clock, Ann knows the following disjunctive proposition: the clock is either pointing at 12:16, 12:17 or 12:18.

First we present Williamson's argument against the KK principle in this context of inexact knowledge (2000: chapter 5). Single agent epistemic modal logic will be used to analyze the argument.[6] We read the modal sentence $Kp_i$ as "Ann knows that the clock is pointing at $12:i$;" $\hat{K}\varphi$ is an abbreviation for $\neg K \neg \varphi$, and translates to "for all Ann knows, $\varphi$." The first premise in the argument against the KK principle is that after looking at the clock Ann knows that the time is not 12:00, i.e.

**P1**: $K\neg p_0$.

The second premise of Williamson's argument encodes Ann's knowledge of her own margin-for-error. By the description of the example, the following should be true: $p_{i+1} \rightarrow \hat{K}p_i$; if the clock is pointing at $12:i+1$, then for all Ann knows (or: Ann cannot rule out that) the clock is pointing at $12:i$. This claim follows from Ann's imperfect eyesight, and Williamson assumes that Ann knows this. Hence the second premise is $K(p_{i+1} \rightarrow \hat{K}p_i)$, or equivalently

**P2**: $K(K\neg p_i \rightarrow \neg p_{i+1})$.[7]

The third premise of Williamson's argument is the KK principle, stated as

[6] Although not relevant to the argument presented here, it can be worthwhile to expand the dynamic analysis I develop in this paper for *multi-agent epistemic logic*, in order to reevaluate a recent attack on common knowledge based on inexact knowledge (Lederman, 2017).

[7] We assume that the implication in **P2** is material.

**P3**:$Kp \to KKp$,

which is assumed for reductio.[8] Assuming multi-premise closure of knowledge,[9] the following derivation holds:

$K\neg p_0$ by **P1**

$KK\neg p_0$ by an instance of **P3**

$K(K\neg p_0 \to \neg p_1)$ by an instance of **P2**

$K\neg p_1$ by closure

...

$K\neg p_i$

$KK\neg p_i$ by an instance of **P3**

$K(K\neg p_i \to \neg p_{i+1})$ by an instance of **P2**

$K\neg p_{i+1}$ by closure

...

$K\neg p_{17}$

The last line is a contradiction, given the assumption that $p_{17}$ is true and the factivity of knowledge. Note how the same reasoning pattern is *iterated* multiple times in the derivation. Williamson's conclusion is that **P3** is false, i.e. the KK principle fails to hold in general.

Williamson's commitment to the second premise of the argument is partly based on his externalist epistemology, which is exemplified with a commitment to the safety condition of knowledge. The latter condition requires that if you know $\varphi$, then you could not have been wrong in very similar cases. Knowledge entails an error free buffer zone. In the unmarked clock example, we take the $i$ case and the $i+1$ case as very similar. Thus, Ann's imperfect eyesight together with the safety condition of knowledge implies that $p_{i+1} \wedge K\neg p_i$ is impossible. Supposing $K\neg p_i$ and that Ann cannot visually discriminate between $i$ and $i+1$, it follows that Ann would have wrongly believed $\neg p_i$ in the very similar case in which $p_i$ was true — contrary to the safety condition. Since $p_{i+1} \wedge K\neg p_i$ is impossible, $p_{i+1} \to \hat{K}p_i$ follows, and since all of this can be concluded by Ann with some reflection, we

[8]The KK principle is sometimes presented in weaker formulations, e.g. with the language of *being in a position to know* instead of knowledge, or with some additional doxastic constraints. For discussions, see, e.g., Greco (2014b: p. 173-174) and Stalnaker (2015: p. 28). These subtleties will not affect the arguments in this paper.

[9]The closure of knowledge will not be the focus of this paper, and will be assumed throughout as an idealization, even if its failure can break Williamson's derivation. See (Williamson 2000: p. 117) for a defense of closure in this context. A different approach would be to assume closure for the concept of *being in a position to know*, but see Yli-Vakkuri and Hawthorne (forthcoming) for complications.

may assume she knows it, hence **P2**.[10] I will later argue for the rejection of **P2** (in Section 4), but my criticism is about Williamson's implementation of the buffer zone intuition (in the form of **P2**), not about the philosophical idea itself.

The concept of safety — as well as the standard semantics of epistemic logic — is modal. It is therefore natural to model the above syntactic argument using Kripke semantics. Recall that $K\varphi$ is true in a world $w$ iff all the worlds related to $w$ by the epistemic indistinguishability relation (denoted with $R$) are worlds in which $\varphi$ is true. Moreover, since knowledge is factive, i.e. the $K$ modality validates the $K\varphi \to \varphi$ axiom, the $R$ relation must be reflexive. Consider the model in Figure 1, where the $R$ relation is represented graphically by the solid arrows:

$$\cdots \longleftrightarrow 15 \longleftrightarrow 16 \longleftrightarrow \underline{17} \longleftrightarrow 18 \longleftrightarrow 19 \longleftrightarrow \cdots$$

Figure 1: Williamson's intended model (reflexive arrows omitted, actual world underlined). The model satisfies **P1** and **P2** but not **P3**.

Note that at $w_{17}$, the world in which $p_{17}$ is true (and the clock is pointing at 12:17), Ann knows $\neg p_0$, i.e. $w_{17} \models K\neg p_0$, as all worlds accessible from $w_{17}$ with the $R$ relation (i.e. $w_{16}, w_{17}, w_{18}$) are $\neg p_0$ worlds. Moreover, Ann knows the disjunction $p_{16} \vee p_{17} \vee p_{18}$ for the same reason. One can also check that $K\neg p_i \to \neg p_{i+1}$ is true at any world in the model: if Ann knows that $p_i$ is not the case, then she is not located at a (very similar) $i+1$ world. Since $K\neg p_i \to \neg p_{i+1}$ is true everywhere, it is known by Ann, validating **P2**. **P3**, the KK principle, is false in the model. Example: in the actual world $w_{17}$, $K\neg p_{15}$ and $\neg KK\neg p_{15}$ are the case. For world $w_{17}$ is accessible to $w_{16}$ in which $\neg K\neg p_{15}$. Figure 1 models Williamson's conclusion of the unmarked clock argument: inexact knowledge without KK.[11]

## 2 Introspection and Inexact observations

How should we model inexact observations? As we work with finite epistemic models and a coarse-grained conception of propositions, each observed proposi-

[10]If **P2** is based on the safety condition of knowledge, then one way of blocking Williamson's argument is by rejecting safety. See Neta and Rohrbaugh (2004) for this direction. But even if the safety condition is false in general, it seems like a very reasonable constraint in cases of inexact observations (Williamson 2008).

[11]I am not claiming that the model in Figure 1 is *the* intended model for the unmarked clock example, rather that it is a good enough simplified model. The model introduces many assumptions that go beyond **P1, P2** and **P3**. Further complications and considerations might be added, (see Williamson 2014) – but this is a good picture to start with.

tion must have a clear boundary. But we want to model observations that lack this phenomenology of exact boundary. How should we do that? This *difficult* modeling question is not answered in Williamson's argument, as the act of observation itself is not part of the model. In Figure 1 we see the epistemic model after observation. But how did we get there?

Similar issues are reflected, to some degree, in Williamson's syntactic representation of his argument against KK. The $K$ operator of static epistemic logic (which Williamson uses in his argument) is ambiguous between the knowledge state *before* looking at the clock and the knowledge state *afterwards*. Static epistemic logic, by itself, cannot draw this distinction. **P1** says that $K\neg p_0$, i.e. that Ann knows, *after looking*, that the minute hand is not pointing at 00. So we conclude that the $K$ operator refers to Ann's knowledge after the observation. **P2** is not as clear. The inner $K$ operator in $K(K\neg p_i \to \neg p_{i+1})$ seems to have the same meaning as the $K$ operator in **P1**: knowledge after the observation. But the outer $K$ operator can be read, prima facie, as representing Ann's knowledge *before* the observation. After all, according to Williamson, knowledge of margin-for-error is obtained by reflection, and this reflection can be done before looking at the clock.

It is crucial for Williamson's argument that at least some of the $K$ operators are understood as knowledge after *the observation.* If a truth-telling Oracle tells Ann that the time is *not* between 12:00 and 12:16, then although **P1** remains true, **P2** becomes false. For then $K\neg p_{16} \wedge p_{17}$ is true, contrary to the margin-for-error principle. This is because in this context the $K$ operator refers to *knowledge after accurate testimony*, and such knowledge state is not governed by a perceptual margin-for-error.[12] **P2** is not set in stone — it is highly dependent on the context of inexact observations. But none of these complications are reflected in the simple language of epistemic logic.

The above remarks, by themselves, do not count as objections to Williamson's argument. Given the argumentative context and the intended model Williamson is using, it is rather clear that according to Williamson, all $K$ operators should be understood as knowledge after the observation. However, a proponent of the argument should allow for this distinction to be made and incorporated into the argument. If the argument is robust, it should easily survive such modification.

So let us make this modification. Start by adding an *update* operator, a familiar addition from *dynamic epistemic logic* or DEL (Baltag & Renne 2016), to the for-

[12]A similar point is made by Sharon and Spectre (2008).

mal language, representing the act of inexact observation.[13] Given any proposition $e$ that can be observed, we add to the language the modal operator $[e]$, representing the observation that $e$. The modal operator $[e]$ is used to describe the result of the experience that $e$ on Ann's epistemic state. As a rough approximation, we can read the formula $[e]\varphi$ as stating "as a result of Ann's veridical experience that $e$, $\varphi$ is the case," or as "if Ann has the veridical experience that $e$, then $\varphi$", where the conditional is not understood as a materiel conditional.[14] By *veridical* we just mean that if we are in a not-$e$ world, then the formula $[e]\varphi$ is vacuously true — our focus is on updating with veridical evidence that possibly generates knowledge.[15] $[e]$, as a 'box' type modal operator, has a 'diamond' dual $\langle e\rangle$, which is equivalent to $\neg[e]\neg$. The only difference between $\langle e\rangle\varphi$ and $[e]\varphi$ is in the treatment of non-veridical $e$: if $e$ is false in the world of evaluation, then $\langle e\rangle\varphi$ is false, while $[e]\varphi$ is vacuously true. Put syntactically, the following is going to be a valid principle: $\langle\psi\rangle\varphi \leftrightarrow [\psi]\varphi \wedge \psi$. For ease of presentation, I will mainly use the box version of the update operator.

Conceptually, a key feature of the $[e]$ update operator is that it is an epistemic, but not necessarily a successful update. It is epistemic because it is used to model change of knowledge given true information (the agent's change of belief is not modeled in this analysis). At the same time, we are not assuming that it is epistemically successful: it is possible that after an update with true $e$, the agent does not come to know $e$. This feature will be important for modeling certain externalist intuitions later. Even in the case where $e$ is true and the agent has the experience that $e$, we don't wish to assume that the agent automatically comes to know $e$, for the external environment can prohibit the agent from coming to know $e$ (say because

[13] Bonnay and Egre (2011) were first to consider the tools of DEL in order to analyze inexact knowledge. Their approach of using a non-standard semantics for the static base epistemic logic is very different from the one I develop here, as they do not use updates to model inexact observations. The very recent work of Baltag and van-Benthem (2018) uses what I call exact updates to analyze Williamson's inexact knowledge. The latter approach is quite different than mine as it does not try to offer an alternative or an explanation to the margin-for-error principle.

[14] Although it is possible, in this paper we will not develop the analysis of the DEL update operator as expressing a non-material conditional. See Icard and Holliday (2017) for an analysis of the relationship between the update operator of DEL and indicative conditionals. Even though the main function of the $[e]$ operator is to describe the effect of observing $e$ on the agent's epistemic state, the formal syntax of DEL allows for expressions like $[e]p$ (where $p$ is an atomic, non-epistemic formula). In those cases, there is clearly no dependency between observing $e$ and $p$, and the conditional reading of the operator is more appropriate.

[15] The debate whether all evidence is veridical is not at issue here. Even if there is false evidence, such evidence does not generate knowledge (at most it can generate false beliefs). Our focus is on knowledge update, so we can safely restrict our attention to veridical evidence and veridical update operators. We care to model knowledge update and not belief revision given any kind of (possibly false) piece of information. That being said, it is technically possible to extend the formal apparatus to accommodate such cases.

the agent is in an epistemically unsafe situation, or the source of information is unreliable in a given situation). Since $[e]$ is not assumed to be successful, it is hard to phrase in ordinary English; we cannot phrase $[e]\varphi$ with an expression like "as a result of learning $e$, $\varphi$ is the case," as the latter assumes that the agent successfully comes to know $e$.

All discussed cases of inexact knowledge in the literature are cases in which the observation happens just *once* — there is no sequence of observations. Syntactically then, we will not nest the update operators inside other update operators. This simplicity leads to a clear way of distinguishing between knowledge *before the observation* and knowledge *after the observation*: every instance of the $K$ operator inside the scope of the update operator represents the latter; all $K$ operators outside the scope of an update operator represent the former. For instance, the first premise of Williamson's argument will be modified to $[e]K\neg p_0$: *after* Ann is making observation $e$, she knows that $\neg p_0$.[16]

Now, given our richer dynamic epistemic language (as opposed to its static fragment), one can locate principles that should be rejected on weak externalist considerations. Consider the following principle:

**Dynamic Introspection (DI):** $[p]Kp \rightarrow K[p]Kp$.

(DI) says that if after the agent has a veridical experience of $p$, the agent knows $p$, then the agent knows (prior to the experience) that after a veridical experience that $p$, the agent will know $p$. Put differently: suppose that $p$ is true, and that once the agent has the experience that $p$, she comes to know $p$. Then the agent knows that a veridical $p$ experience generates knowledge. It seems that very weak forms of externalism generate counterexamples to (DI). After all, even if a piece of evidence $p$ is true, it does not follow that it is also received safely (or reliably, or by the right causal connections). Consider the following example, formulated with a generic reliabilist language:

**The Tree:** There is a tree next to Bob. At $t_1$, Bob is having a veridical experience of a tree, and as a matter of fact, Bob's vision is reliable. Therefore, Bob knows that there is a tree in front of him at $t_1$. However, Bob does not know that his vision is reliable, and at $t_0$ he cannot conclude that the experience of a tree will result in him knowing that there is a tree in front of him. For all Bob knows, the

[16] Alternatively, one could use the framework of *epistemic temporal logic* (instead of dynamic epistemic logic) to represent the knowledge stages before and after the update with two distinct knowledge operators: $K_0$ and $K_1$. See van-Benthem et al. (2009) for details about the relationship between the two frameworks.

experience of a tree might not be generated in a reliable fashion.

Note then that according to the example, we have (1) $[tree]K(tree)$ — after the veridical experience of $tree$, Bob knows $tree$, and (2) $\neg K[tree]K(tree)$ — Bob does not know, before the experience, that having the experience of $tree$ will result in knowledge of $tree$, since for all Bob knows, his perceptual faculties are unreliable. The conjunction of (1) and (2) provides a counterexample to (DI). I therefore take it as uncontroversial that generic forms of externalism are committed to the failure of (DI).

It is worth mentioning that some internalists and skeptics can be understood as endorsing a form of dynamic introspection. The contra-positive form of (DI) is $\neg K[p]Kp \rightarrow \neg[p]Kp$. The latter formula expresses the idea that if the agent does not have the prior knowledge that their source of information is reliable ($\neg K[p]Kp$), then the agent cannot gain knowledge from that source of information ($\neg[p]Kp$). This constraint seems to capture the internalist objection to externalist epistemology: one cannot gain knowledge from a source that is not known to be reliable. Likewise, the contra-positive form of (DI) can be used to express a standard skeptical argument about perception: you cannot verify that your perceptual capacities are reliable (skeptical assumption), if you cannot verify that, then you cannot obtain knowledge by perception (contra-positive of DI), therefore you do not have perceptual knowledge (skeptical conclusion). These considerations further corroborate my assumption that externalists should be understood as *rejecting* (DI).[17]

It is also worth remarking that (DI), like the KK principle, is a kind of introspection principle, where the consequent iterates a $K$ operator on a condition in the antecedent. But (DI) and KK have different logical origins, as the former is dynamic while the latter is static. As a matter of fact, (DI) is a validity in standard forms of DEL, even if the underlying static epistemic logic is the modal logic $T$, i.e. a logic that invalidates the KK principle. In the logical picture that I am going to present, (DI) is false, while the KK principle remains true. Thus, the two principles are logically independent.[18]

All of this is related to Williamson's argument. I will now show that (DI) is incompatible with a plausible *dynamic* presentation of the unmarked clock example. Thus, Williamson's inexact knowledge argument could be interpreted as a reductio

[17]For a more systematic discussion about dynamic introspection, externalism, and skepticism see Cohen (2020a).

[18]It is thus also an independent question whether the Tree example is compatible with the KK principle. I leave this question aside. My point is that the example is clearly not compatible with Dynamic Introspection.

argument against (DI), once Williamson's premises are understood dynamically.

The problematic step in the Williamsonian derivation from the introductory section is the move from Ann's knowledge that $\neg p_{15}$ to her 'knowledge' that $\neg p_{16}$. The latter step is defective as we assumed that Ann knows $p_{16} \vee p_{17} \vee p_{18}$, but not more. In the derivation, we make this step by using $K(K\neg p_{15} \to \neg p_{16})$, an instance of **P2**. With the update operators, I will attempt to explain where the argument goes wrong. What follows is my reconstruction of the problematic part of the unmarked clock argument.

Fix $e$ to be the proposition $p_{16} \vee p_{17} \vee p_{18}$. My **main assumption** is $[e]Ke$: after observing $e$, Ann knows $e$. It is an analytic truth that $e \to \neg p_{15}$, so we can assume that it is known by Ann before and after the observation, and known to be so. We thus have $K[e]K(e \to \neg p_{15})$; call this my auxiliary assumption. Given the closure of knowledge, the closure (or distribution) of the update operator, and our main and auxiliary assumptions we can deduce

**P1′** : $[e]K\neg p_{15}$

(see the appendix for the full deduction). This is the first premise in *our* reconstruction of Williamson's argument. Next, we wish to translate Williamson's knowledge of margin-for-error assumption (his premise **P2**) into the dynamic language. I propose the following:

**P2′** : $K[e](K\neg p_{15} \to \neg p_{16})$.

This is the dynamic version of Williamson's **P2**: Ann has the prior knowledge that after observing $e$, it is going to be the case that: if she knows $\neg p_{15}$, then $p_{16}$ cannot be true, given her margin-for-error. For if the evidence $e$ is strong enough to generate knowledge that excludes the $p_{15}$ possibility, then, given Ann's observational margin-for-error, it could not have been generated in world $w_{16}$. Note that the addition of the dynamic operator allows us to distinguish knowledge before and after the observation. The inner $K$ is scoped by $[e]$, thus representing knowledge after the observation that $e$; the outer $K$ is not scoped by the operator, representing knowledge attained before the observation. **P2′** gives us a more fine-grained description of the situation at hand.

Recall that we assumed that the strongest proposition about the position of the clock that Ann learned from the observation is $e$. In Williamson's original derivation, we can derive $K\neg p_{16}$, which implies that Ann knows more than $e$, a contradiction to what we assume. Similarly, with our assumptions we can derive that Ann knows that after observing $e$, $\neg p_{16}$ is the case. This contradicts what we assumed, as we assumed that Ann does not know that $\neg p_{16}$ after observing $e$ (we started with the assumption that the strongest proposition known to Ann purely about the position of the clock after the observation is $p_{16} \vee p_{17} \vee p_{18}$). Thus, the **main assumption**, premises **P1**$'$, **P2**$'$ and (DI) lead to an absurdity, and I conclude that (DI) should be rejected.

I reserve the full formal derivation to the Appendix, and sketch it here: the **Main assumption** states that $[e]Ke$; together with (DI) and the auxiliary assumption, it follows that $K[e]K\neg p_{15}$. **P2**$'$ states that $K[e](K\neg p_{15} \to \neg p_{16})$; under the assumption that the update operator distributes over implication, it follows that $K([e]K\neg p_{15} \to [e]\neg p_{16})$. By the closure condition of knowledge, it follows that $K[e]\neg p_{16}$ from the above two conclusions. In the form of a derivation:

| | |
|---|---|
| (1) $[e]Ke$ | **Main assumption** |
| (2) $K[e]Ke$ | by (DI) on (1) |
| (3) $K[e]K(e \to \neg p_{15})$ | auxiliary assumption |
| (4) $K[e]K\neg p_{15}$ | closure (see Appendix) |
| (5) $K[e](K\neg p_{15} \to \neg p_{16})$. | **P2**$'$ |
| (6) $K[e]\neg p_{16}$ | from (2) and (3), by closure (see Appendix). |

Line (6) contradicts the assumption that the strongest proposition that Ann learns from observing $e$ is $e$, as it states that Ann knows that observing $e$ implies something stronger, namely $\neg p_{16}$. Crucially, note that we have derived a conclusion contrary to our initial assumption *without* the use of the KK principle.

The most controversial principle in the above paragraph is (DI). Much less controversial is the assumption that simple analytic truths like $e \to \neg p_{15}$ are known, and known to remain known after any veridical observation (i.e. $K[e]K(e \to \neg p_{15})$, the auxiliary assumption). We also assume that the update operator $[\cdot]$ distributes over implication: if after the $\varphi$ update, $\alpha \to \beta$ is true, and after the $\varphi$ update, $\alpha$ is true, then after the $\varphi$ update, $\beta$ is true. Assuming weak externalist tendencies, which are in tension with (DI) anyway, rejecting (DI) seems like the most plausible response.

Upshot: Williamson's formal argument does not distinguish between knowledge before and after the inexact observation. Once this distinction is made, we can enrich Williamson's argument with dynamic operators. When we do so, we see that the dynamic premises conflict with the dynamic principle I called (DI). This conflict offers an explanation to the tension between inexact observations and introspection. This dynamic explanation is different from Williamson's static explanation, as it involves different types of introspection principles. Note, however, that the alternative dynamic explanation is by itself not in conflict with Williamson's static argument. After all, the dynamic language is an *extension* of the static language; the two are perfectly compatible. It is hence fair to ask whether Williamson's static premises are true after the act of observation, and so whether inexact observations are in conflict with *both* static and dynamic introspection. In section 4 I will return to this question. Before I do that, I offer an account of inexact observations as a special kind of update, an *inexact update*.

## 3 A Semantic Perspective

While the last section has focused on the syntactic argument, here I present my novel semantics for inexact observation. The goal is to show that the KK principle and formulas like the **main assumption**, **P1**$'$ and **P2**$'$ are mutually compatible with the failure of (DI). A natural way to argue for compatibility of a set of assumptions is by invoking a model, which is what I do here. The model will also explain my novel approach to inexact observations as *inexact updates*.

The main challenge is to model the epistemic effect of inexact observations. To do so, we work with two models: the *initial model* (the situation before the observation has taken place), and the *updated model* (the situation after the observation). The semantic clause of the update operator will tell us how to compute the updated model from the initial model.

The semantics of the $K$ operator remains the same: $K\varphi$ is true iff $\varphi$ is true at all epistemically accessible worlds. The basic idea behind the semantics of the update operators in DEL is the familiar Stalnakerian notion of update: updating with proposition $P$ has the effect of eliminating all the not-$P$ worlds from the initial model. This is a good enough semantics for modeling exact updates, like learning from reliable testimony, but we will need to tweak it to accommodate inexactness.

I assume that in the initial model (before looking), Ann does not know anything about the position of the minute hand of the clock: $\neg Kp_i$, for any $p_i$. The principle that $Kp_i \to KKp_i$ thus follows vacuously. Moreover, we can assume that Ann knows that she does not know the position of the minute hand. Ann has no reason to think that she knows the time, and no deception is assumed in the example: hence $K\neg Kp_i$ holds for any $i$, and so $\neg Kp_i \to K\neg Kp_i$ follows as well. In order to capture these assumptions, we can let the initial model be an $S5$ model in which all worlds are epistemically connected: $Rw_iw_j$ for any $i$ and $j$.[19]

[19]Since the initial the model is an S5 model, positive and negative introspection hold for *any* $\varphi$, not just for the $p_i$'s. Since what we care about is Ann's knowledge of the clock (i.e. about $p_i$), this

Now, we add to the initial model another relation, the *perceptual inexactness* relation $P$, that intuitively specifies which worlds will be perceptually indistinguishable *during* an observation.[20] Unlike Ann's initial epistemic state, perceptual indistinguishability depends on which state is actual. If the actual state is $i$ then, on observation, Ann cannot perceptually distinguish it from $i+1$ and $i-1$, according to the informal story about her margin-for-error. In other words, we have that for all $w_i$, $Pw_iw_{i+1}$ and $Pw_iw_{i-1}$.[21] Moreover, as an indistinguishability relation, $P$ is assumed to be reflexive. The *initial* epistemic model is depicted in Figure 2.

$$\cdots \dashleftarrow\dashrightarrow 15 \dashleftarrow\dashrightarrow 16 \dashleftarrow\dashrightarrow \underline{17} \dashleftarrow\dashrightarrow 18 \dashleftarrow\dashrightarrow 19 \dashleftarrow\dashrightarrow \cdots$$

Figure 2: The initial model $M$ according to my story (before Ann looks at the clock). The epistemic indistinguishability relation is the universal S5 relation, which is not depicted. The dashed arrows represent the relation $P$, of perceptual inexactness.

The model in Figure 2, unlike that of Figure 1, contains two relations: the universal epistemic $R$ relation, connecting all worlds together (not depicted in the figure), and the perceptual inexactness relation $P$, depicted with the dashed lines.

What happens to the initial model when Ann has the $e$ $(= p_{16} \vee p_{17} \vee p_{18})$ experience when looking at the clock at the actual world $w_{17}$? The idea is that we eliminate all the not-$e$ worlds from the model, unless these worlds are perceptually indistinguishable from $w_{17}$ (according to $P$). In general, the updated model, resulting from observing $\varphi$ at $w$, has the following key property: all worlds in it are either (1) worlds in which $\varphi$ is true, or (2) worlds that are perceptually indistinguishable from $w$. Put formally, the semantic satisfaction clause for the update operator is:

- $M,w \models [\varphi]\psi \;\Leftrightarrow\;$ if $M,w \models \varphi$ then $M_{\varphi,w},w \models \psi$.

The antecedent on the right hand side of the equivalence guarantees that when $\varphi$ is false, the expression $[\varphi]\psi$ is vacuously true. The consequent checks that $\psi$ is true in the updated model $M_{\varphi,w}$, which is formally defined as $M_{\varphi,w} = (W',R',P',V')$:

$W' = \{v \in W \mid Pwv \;\; OR \;\; M,v \models \varphi\} = \{v \in W \mid Pwv\} \cup \{v \in W \mid M,v \models \varphi\}$

$R' = R \cap W'^2$

$P' = P \cap W'^2$

idealization seems harmless.

[20] See Halpern (2008) for different approach that uses two relations in order to analyze related cases of perpetual vagueness. Dutant (2007) builds on Halpern's approach and provides an infallibilist critique of Williamson's margin-for-error argument.

[21] We could complicate the structure of the $P$ relation to allow for varying margin-for-error the same way Williamson is varying his $R$ relation in (Williamson 2014).

$V' = V$.

$W'$, the set of worlds obtained by observing $\varphi$ at $w$, is just the *union* of the $\varphi$ worlds with the perceptually indistinguishable worlds. Note that if we wish to model the special case in which updates are exact, we just need to set the relation $P$ s.t. it does not connect any two distinct worlds (i.e. as empty apart from being reflexive). On such frames, the update operator behaves the same as in standard DEL.[22]

The novelty of the above semantics lies in the fact that updating with the same $\varphi$ at different worlds results in different updated models.[23] The leading motivation behind the semantics is based on a familiar intuition in epistemology: in the *good case*, veridical evidence generates more knowledge than in the *bad case*. Read, for instance, the *good case* as the case where our perceptual capacities are reliable, and the *bad case* as the case where our capacities are not as reliable. Further assume that $\varphi$ is a true piece of information ($\varphi$ is true at the world of evaluation). In the good case, observing $\varphi$ leads to knowing $\varphi$. In the bad case, observing $\varphi$ leads to knowing a weaker proposition, say $\psi$. In the worst case imaginable, i.e. the extreme skepticism scenario, the veridical $\varphi$ experience leads to no new knowledge at all (i.e. $\psi$ is $\top$).

Let's consider the concrete example of the unmarked clock: updating with $e$ at $w_{17}$. The result is in Figure 3:

$$16 \longleftrightarrow 17 \longleftrightarrow 18$$

Figure 3: $M_{e,w_{17}}$ The result of updating with $e$ (= $p_{16} \vee p_{17} \vee p_{18}$) at world 17.

Compare that to updating with the same $e$ at world $w_{16}$:

$$15 \longleftrightarrow 16 \longleftrightarrow 17 \longleftrightarrow 18$$

Figure 4: $M_{e,w_{16}}$, updating with $e$ (= $p_{16} \vee p_{17} \vee p_{18}$) at world 16.

and at world $w_{18}$:

$$16 \longleftrightarrow 17 \longleftrightarrow 18 \longleftrightarrow 19$$

Figure 5: $M_{e,w_{18}}$, updating with $e$ (= $p_{16} \vee p_{17} \vee p_{18}$) at world 18.

[22]In Cohen (2020b), I develop a more general logic for inexact and opaque updates, which can also be used to model the example analyzed here.

[23]Note that both standard DEL and standard Bayesian update lack this property. Such updates are insensitive to the world of evaluation.

When we update with $e$ at world $w_{16}$, we cannot eliminate the close world $w_{15}$ from the updated model. This is why the model in Figure 4 contains world $w_{15}$. Similarly, world $w_{19}$ remains in the updated model if we update with $e$ at world $w_{18}$ (in Figure 5). The key issue is rather trivial: inexact updates cannot eliminate close worlds, and the set of worlds that count as close changes according to the world of evaluation.

In the above models, relative to observation $e$, $w_{17}$ is the good case while $w_{16}$ and $w_{18}$ are the bad cases (which are not, however, skeptical cases). In $w_{17}$, having the veridical experience of $e$ results in knowing that $e$. To see why, note that in Figure 3, Ann knows $e$, as $e$ is true everywhere in that model. Formally: $M, w_{17} \models [e]Ke$, since $M, w_{17} \models e$ and $M_{e,w_{17}}, w_{17} \models Ke$. In the not so good case, $w_{16}$, having the $e$ experience does not result in knowing $e$, but the weaker $e \vee p_{15}$. Formally: $M, w_{16} \models [e]K(e \vee p_{15})$ (consult the updated model in Figure 4 to see why). Similarly, $M, w_{18} \models [e]K(e \vee p_{19})$. In the good case, Ann gets the most out of the evidence — in the other cases, she gets less.

Recall that the actual world in the example is indeed $w_{17}$. Thus, since $M, w_{17} \models [e]Ke$, the model satisfies what I have previously called the **main assumption**. Clearly, **P1**$'$, $[e]K\neg p_{15}$ holds as well in $w_{17}$. Consider then the truth of **P2**$'$: $M, w_{17} \models K[e](K\neg p_{15} \to \neg p_{16})$. The only world that can witness the falsity of the known conditional is $w_{16}$, in which the consequent is false. The question is then whether $M, w_{16} \models [e](K\neg p_{15} \to \neg p_{16})$ holds. The answer is yes, because the antecedent is false: at world $w_{16}$, having the $e$ experience does not result in knowing $\neg p_{15}$ (consult Figure 4). Thus, **P2**$'$ is true in my model; the agent knows the epistemic effects of her margin-for-error. Moreover, the $KK$ principle holds in all the models of Figures 2 to 5, as in these models the epistemic $R$ relation is universal.[24] Static introspection is not a cause for concern.

Finally, and most importantly, note that (DI) fails in my model. We have that $M, w_{17} \models [e]Ke$, so according to the latter principle we should also have $K[e]Ke$ at $w_{17}$. But that is false, since, from the perspective of the initial model, Ann considers $w_{16}$ to be epistemically possible. In $w_{16}$ however, $\neg[e]Ke$ holds, thus $M, w_{17} \models \neg K[e]Ke$.

The above line of reasoning is (again) rather familiar within externalist epistemology: although Ann is actually in the good case ($w_{17}$), and although the observation of $e$ results in knowledge in the good case, *for all Ann knows*, she is in the bad

[24]One can also construct updated models which are not S5 models, rather only S4. The 5 axioms does not play any crucial rule in my argument.

case $w_{16}$, and in the bad case, the same experience will not result in the same state of knowledge. Since Ann does not initially know whether she is in the good case or not, she cannot initially know how the $e$ experience will affect her knowledge state. Ann does not have dynamic introspection.

This completes the basic semantic picture of inexact observations. The model I offered (Figure 2) shows that it is possible to satisfy the KK principle and prior knowledge of margin-for-error, while falsifying dynamic introspection. However, the model does not assume that the agent has unrestricted knowledge of margin-for-error posterior to the observation. This is because I reject the claim that agents have unrestricted knowledge of their margin-for-error both prior and posterior to observations. I explain this in the next section.

## 4 Back to Safety

I have offered an alternative explanation to the tension between inexact observations and introspection, based on the language of dynamic epistemic logic and my novel semantics of inexact updates. In particular, I reformulated Williamson's margin-for-error principle with inexact updates and argued that it fits naturally with the commitments of externalist theories of knowledge. The fact remains that Williamson's three static premises are inconsistent in (static) epistemic logic, so I must reject one of them. I reject **P2**, Williamson's formulation of the agent's knowledge of their own margin-for-error. In this section I explain how knowledge of margin-for-error is compatible with a rejection of **P2**. I argue that Williamson's reasoning pattern (presented in the derivation in Section 1.1) cannot be iterated. Even in cases where the agent knows her margin-for-error, this knowledge can be only used once. The objection I present here is in the same spirit of earlier critics of Williamson's argument, notably Sharon and Spectre (2008) and Dokic and Égré (2009), but my overall dynamic analysis allows to present this type of criticism in a novel, more comprehensive perspective.

Up to now, we have formulated knowledge of margin-for-error (of one unit) as $K[e](K\neg p_i \rightarrow \neg p_{i+1})$. This formula describes the prior *de-dicto* knowledge Ann has about the effect of making an observation $e$ with a margin-for-error of 1 unit. This prior de-dicto knowledge does not imply that posterior to the observation event, Ann has the *de-re* knowledge that *that* event was an inexact observation event with a margin-for-error of 1 unit. After all, Ann can be uncertain as to her exact margin-for-error at that particular observation event. To avoid this complication, we can simplify things by considering a margin-for-error of an even smaller value. Consequently, we can assume that Ann can know that the observation event she experienced followed a margin-for-error of that small unit. So let us assume that Ann is in a position to know that, posterior to the observation event, her perceptual margin-for-error for that event must *at least* be 0.1. Thus, we assume that the formula $[e]K(K\neg p_i \to \neg p_{1+0.1})$ is true.[25]

According to Williamson, if the KK principle is true, then after the observation, Ann can use her margin-for-error knowledge again and again to rule out $p_{16}$, $p_{16.1}, p_{16.2}$... etc. I agree that Ann can use her margin-for-error once to rule out the $p_{16}$ possibility, but Williamson is wrong, I argue, in assuming that this reasoning process can be iterated. After the observation *and* Ann's reasoning process Ann considers it possible that she knows that it is not $p_{16}$ and that the world is actually 16.1. In other symbols, $\hat{K}(K\neg p_{16} \wedge p_{16.1})$ is the case. This is a counterexample to Williamson's **P2** for margin-for-error 0.1. The sentence $\hat{K}(K\neg p_{16} \wedge p_{16.1})$ says that Ann considers it possible that she is on the 'edge', knowing $\neg p_{16}$ very close to $p_{16}$ (in a world where $p_{16.1}$ is true). However, this conclusion should not count as violating the margin-for-error principle and the safety intuition behind it, since the margin-for-error principle is meant to describe the agent's knowledge state *resulting from* an inexact observation, *not* her knowledge state in general. Recall that an inexact observation is safe iff as a result of the observation the actual world is surrounded by a large enough buffer zone of close possible worlds. In this sense, Ann's knowledge gained directly by the inexact observation of $e$ is safe. This knowledge state is then combined with Ann's background knowledge about her own margin-for-error, resulting in a new knowledge state in which the 0.1 subintervals on the sides have been ruled out. Now, this new knowledge state should not be constrained by observational inexactness anymore, as it is not purely observational knowledge at that point. The margin-for-error principle only applies to the knowledge state obtained as a result of an inexact observation, it does *not* apply to knowledge states obtained by observation *together* with other non-observational means. The margin-for-error principle is a principle about *perceptual* knowledge; it is not a principle about knowledge *in general*.

To emphasize this point, consider the following variation of the unmarked clock example. Suppose that before looking at the clock, Ann does not form any beliefs about her margin-for-error. She looks at the clock and comes to know that the

[25]Nothing hinges on this choice of values. For whatever value of margin-for-error we choose, we cannot iterate the Williamsonian reasoning pattern from section 1.1.

minute hand is somewhere in the interval 16-18; suppose it actually points to 17 and that Ann's actual margin-for-error is 1. Now Ann's optometrist shows up and tells her that given the conditions of *the observation she just made*, she cannot reliably perceptually distinguish 0.1 distance on the clock face: it is the case that if the minute hand points to $p_i$, then for all Ann knows it points to $p_{i\pm 0.1}$. In other words, the optometrist tells Ann that her margin-for-error is *at least* 0.1. Since Ann's optometrist is a known reliable source (let's assume so), Ann comes to know that her margin-for-error is at least 0.1. Ann then uses her knowledge gained from the optometrist (and the KK principle) to cut her uncertainty interval by 0.1 on both sides, coming to know that the minute hand is somewhere in the interval 16.1-17.9. More specifically, we assume that Ann knows that the clock is not pointing to 15.9 ($K\neg p_{15.9}$) as a result of the inexact observation, and that she knows that if she knows that it does not point to 15.9 it cannot be pointing to 16 ($K(K\neg p_{15.9} \rightarrow \neg p_{16})$), by the optometrist testimony. By the KK principle, $KK\neg p_{15.9}$ obtains, and so by the assumed closure of knowledge, it follows that $K\neg p_{16}$. Likewise, from the assumptions that $K\neg p_{18.1}$ (due to the inexact observation) and $K(K\neg p_{18.1} \rightarrow \neg p_{18})$ (due to the testimony), it follows that $K\neg p_{18}$. The remaining epistemic possibilities range from 16.1 to 17.9.

Importantly, note that Ann cannot reuse the knowledge she gained from her optometrist to conclude anything stronger; she cannot iterate the process. The optometrist did not convey the information that $p_{i+0.1} \rightarrow \hat{K}p_i$ is true in general; they conveyed the information that *after the observation Ann just made* it is true that $p_{i+0.1} \rightarrow \hat{K}p_i$. Once Ann uses the information she got from the optometrist, the sentence $p_{i+0.1} \rightarrow \hat{K}p_i$ becomes false, because the context has changed: now the $K$ operator does not refer solely to knowledge after the observation, but to knowledge after making an observation and learning from testimony. The optometrist did *not* say "after anything you learn, it must be the case that $p_{i+0.1} \rightarrow \hat{K}p_i$", they said "after you make an inexact observation, it must be the case that $p_{i+0.1} \rightarrow \hat{K}p_i$." The latter statement is a true description of the effect of inexact observation. The former statement is a false and ungrounded description of Ann's general knowledge structure. Surely it is possible for Ann to come to know $K\neg p_i \wedge p_{i+0.1}$ by some other non-observational means.

Williamson thinks that there is something problematic about a situation in which Ann considers it possible that the clock points to 16.1 and at the same time she knows that is does not point to 16. Williamson believes that such a situation is in a direct conflict with the safety condition of knowledge. I disagree. Given

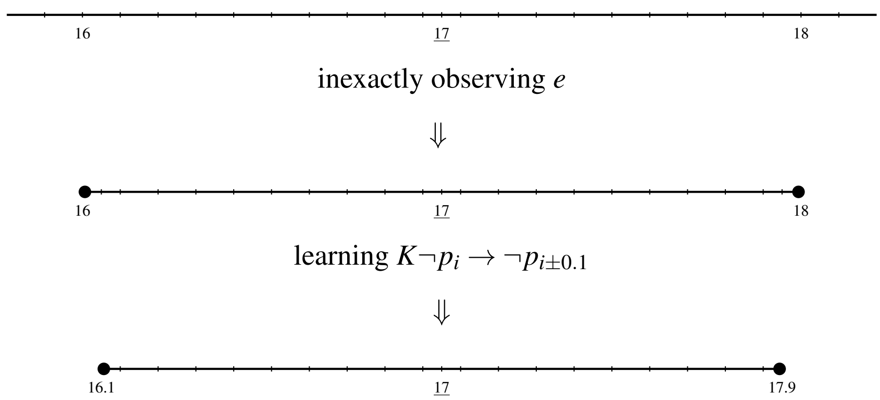


Figure 6: Two model transitions: Ann first inexactly observes $e$, then incorporates her margin-for-error knowledge.

the fine-grained dynamic analysis I propose, this situation can be explained. Consider the optometrist version of our story again. Ann starts by making an inexact observation and learning that the minute hand is between 16 and 18. After the optometrist informs Ann about her perceptual margin-for-error, she concludes that the minute hand must be between 16.1 and 17.9. When asked whether she thinks it is possible that the minute hand is in fact pointing to 16.1, she can respond: "for all I know, the minute hand is actually pointing to 16.1. I can tell that it is not pointing to 16 because this contradicts the observation I made together with the optometrist's information. But I cannot conclude anything stronger. In particular, it is possible, as far as I can tell, that I have initially observed that the minute hand is pointing somewhere between 16 to 18 and that it was actually pointing to 16.1. This state of affairs does not contradict my assumption that the optometrist spoke truly (i.e. the margin-for-error principle is correct). If I knew that my perceptual margin-for-error is larger than 0.1, I could have ruled out the possibility that the clock is actually pointing to 16.1. But I don't know that." In this context, I think there is nothing problematic about Ann's response. Ann's perceptual knowledge, obtained by an inexact observation, is safe. Ann's resulting knowledge state, after taking into account the optometrist's testimony, does not violate the safety requirement of perceptual knowledge (i.e. the margin-for-error principle), because it is not purely a perceptual knowledge state anymore.

Figure 6 graphically summarizes my analysis of the situation, incorporating the lessons from this section and section 3. It contains three (simplified) epistemic models and two updates: the top model represents Ann's initial epistemic state, before making any observation.[26] The middle model represents Ann's epistemic state after inexactly observing $e$ but before learning about her margin-for-error. The transition between the top model and the middle model was explained in detail in section 3. The bottom model represents Ann's knowledge state after learning from the optometrist about her margin-for-error. It is obtained by eliminating all the possible worlds in the middle model in which $K\neg p_i \rightarrow \neg p_{i\pm 0.1}$ is false (i.e. by an exact update with $K\neg p_i \rightarrow \neg p_{i\pm 0.1}$). Note that it is only worlds $w_{16}$ and $w_{18}$ which are eliminated. In $w_{16}$, for instance, we have $K\neg p_{15.9} \wedge p_{16}$, contradicting the information the optometrist conveyed, so it is eliminated. The important thing to note is that when the optometrist announces that $K\neg p_i \rightarrow \neg p_{i\pm 0.1}$, they refer to the knowledge state in the middle model, the knowledge resulting from an inexact observation, not to any other knowledge state.[27] In world $w_{17}$ of the bottom model, it is true that $\hat{K}(K\neg p_{16} \wedge p_{16.1})$. As I explained in the previous paragraph, this is not in conflict with the safety condition for perceptual knowledge. In the bottom model, the $K$ is interpreted as the knowledge state obtained from a combination of an inexact observation and the information received from the optometrist. Therefore, the safety condition for perceptual knowledge (i.e. the margin-for-error principle) does not apply to this state.

The addition of the optometrist is not essential to my account. The only difference between the two versions of the story is that in Williamson's story, knowledge of margin-for-error is obtained before the observation (by reflection); in the optometrist story, this knowledge is obtained after the observation (by testimony). The stage (and method) in which the agent learns their margin-for-error should not affect their final knowledge state. I believe, however, that my modified version of the story makes it easier to recognize that the margin-for-error principle is true only for the knowledge state obtained by the inexact observation. Figure 6 can equivalently be interpreted as representing the process where Ann first observes $e$ and then incorporates her background knowledge about her margin-for-error. Under my account, Ann can have knowledge of margin-for-error posterior to the observation. She can then use that knowledge once, but not more than that. Figure 6 represents this two step process of first making an inexact observation, and then incorporating one's knowledge of margin-for-error.

[26] The models are simplified because the $P$ relation is not drawn.

[27] In the terminology of dynamic epistemic logic, the announcement made by the optometrist is not *successful*, because it is not known after the announcement (see Baltag and Renne 2016). This is an indication that the content of the announcement is context sensitive, in the sense used within dynamic epistemic logic (see, e.g. Holliday 2018).

One might offer the following objection to my analysis: since Ann's final knowledge state is the result of both her inexact observation and her knowledge of her margin-for-error, her final knowledge state is inexact as well. And if her final knowledge state is inexact, then it must follow a margin-for-error principle by itself. If Ann knows this margin-for-error (and there is no reason to assume otherwise), she can apply it again, come to know a stronger proposition, repeat the same reasoning again, and so potentially reach Williamson's contradictory conclusion.

There are several problems with the objection. First, the objection wrongly assumes that any knowledge state that is partially the result of an inexact observation must itself be an inexact knowledge state, and so follow a margin-for-error. Here is a counterexample to this assumption. Consider the following scenario: Ann makes an inexact observation and comes to know that the minute hand is not pointing to zero, $\neg p_0$. This is an assumption that both Williamson's story and my story can accommodate. Later, an Oracle tells Ann that the clock is not pointing in the range $p_1 - p_{16}$, nor does it point in the range $p_{18} - p_{59}$ (leaving only $p_0$ and $p_{17}$ possible). After receiving the Oracle's information, Ann comes to know $p_{17}$, using her knowledge from the inexact observation, the Oracle's information, and her deductive abilities. Thus, Ann knows exactly where the minute hand is pointing to, and this knowledge state is the result of both the inexact observation and the Oracle's information. According to the assumption in the objection, Ann's final knowledge state is inexact, because it is partially the result of an inexact observation. If so, then her final knowledge state must follow a margin-for-error. But Ann's final knowledge seems to be exact in this scenario (she knows exactly where the minute hand is pointing), and so her knowledge does not follow a margin-for-error anymore. Thus, the scenario offers a counterexample to the claim that every knowledge state partially obtained by an inexact observation must be an inexact knowledge state that follows a margin-for-error. It is possible to reach a knowledge state that does not follow a margin-for-error from a previous knowledge state that does.

Second, even if we assume that the agent's final knowledge state is inexact, it is unclear how this leads to a contradiction. For the sake of the argument, grant the objector the assumption that—unlike the scenario described in the last paragraph—every method the agent has for obtaining knowledge is in some way inexact, and so follows some margin-for-error. Even with this assumption, it is far from clear that Williamson's neat contradictory derivation follows. The force of Williamson's argument comes from the fact that the perceptual margin-for-error principle $p_i \rightarrow \hat{K} p_{i+\varepsilon}$ is so intuitive: clearly, there must be some $\varepsilon$ such that I cannot visually discriminate between the minute hand being in position $i$ ($p_i$) and position $i+\varepsilon$ ($p_{i+\varepsilon}$). Other margin-for-error principles (resulting from the agent's non-visual inexact methods of gaining knowledge) will not be so easy to accept, or even to formulate. Assume that Ann's deductive abilities are inexact, and so the knowledge Ann obtains by deduction follows some margin-for-error. To articulate such margin-for-error, one would first have to come up with a notion of similar possibilities for the outcomes of Ann's deductive inferences, and then formulate a margin-for-error principle based on that notion of similarity. There is no reason to assume that the latter notion of similarity will have anything to do with a notion of similarity based on metric distance (which we use for perceptual inexactness). The same can be said of other potential inexact methods of gaining knowledge, like inexact memory or inexact (i.e. potentially unsafe) testimony. Even if Ann's knowledge state is governed by a further margin-for-error principle(s), there is no reason to assume that those margin-for-error have the form $p_i \to \hat{K}p_{i+\varepsilon}$, which is crucial for Williamson's derivation of contradiction. It is the burden of the objector to offer a compelling story as to why a different margin-for-error principle resulting from another inexact method of gaining knowledge leads to a contradiction. And my response to such attempt will be similar to my earlier response: a margin-for-error principle is applicable for a particular epistemic state; once the margin for error is used by the agent, the epistemic state has changed, and there is no reason to assume that the same margin-for-error applies in the new state as well (a different margin-for-error might apply to the new situation, but there is no contradiction in that).[28]

Williamson's static formulation of knowledge of margin-for-error, $K(K\neg p_i \to \neg p_{i+\varepsilon})$, is inadequate exactly because it does not capture the idea that the inner $K$ in it refers to knowledge *after an inexact observation*. By reusing the margin-for-error principle in his derivation (presented in Section 1.1), Williamson implicitly assumes that the inner $K$ in the principle describes a general knowledge state of the agent. As the optometrist story meant to convey, this is a mistaken assumption. My dynamic formulation of inexact perceptual knowledge fares better. Syntactically, the ability to scope the $K$ in the inexact observation operator $[e]$ allows us

[28] My response essentially appeals to a quantifier shift fallacy. To get a contradictory derivation, the objector needs to assume that there is one type of margin-for-error principle for every way of gaining inexact knowledge. I argue that for every way of gaining inexact knowledge, there is some type of margin-for-error principle. Since I see no reason to assume that the different margin-for-errors have the same structure, I don't see how one type of margin-for-error can be repeatedly applied to obtain a contradiction.

to represent the knowledge which results from an inexact observation. Semantically, the mechanics of my inexact updates allow me to formally connect the margin-for-error principle with the epistemic *result* of an inexact observation. As a consequence, we get a broader picture of *inexact observational knowledge*, as the knowledge that results from inexact observations. I conclude that my dynamic story does a better job in capturing the epistemic aspects of inexact observations.

As I mentioned earlier, the objection that the margin-for-error principle is context specific and should not be assumed in an unrestricted form already appears in the literature (Sharon and Spectre 2008, Dokic and Égré 2009). For this reason, I would like to stress the main differences between my approach and earlier criticisms. First, unlike Dokic and Égré (2009) my analysis does *not* rely on distinguishing between two types of knowledge operators, perceptual and reflective (see also Halpern 2008, Égré and Bonnay 2008, Sharon and Spectre 2008 for similar proposals). My conceptual, syntactic and semantic treatment of the knowledge operator is uniform. My dynamic analysis, however, does allow me to distinguish between the behavior of different sources of information, and to associate different safety requirements with different sources. Moreover, and unlike earlier criticisms, my formal framework cannot be accused of being *ad-hoc*, given that dynamic epistemic logic is an independently motivated formal framework.[29]

Second, and more importantly, unlike earlier criticisms, my analysis manages to capture the compelling elements of Williamson's story. Like Williamson, I believe that there is a tension between inexact observations and introspection, and that this tension is worthy of a philosophical explanation. Furthermore, I believe that Williamson's argument is persuasive *because* it offers an explanation to this tension. Earlier objections have pointed out the flaws in assuming that the margin-for-error principle holds unrestrictedly, but they have not offered an alternative positive account to explain in what ways inexactness is incompatible with introspection. The dynamic account I present in this paper shows how inexact observations are incompatible with a dynamic form of introspection. The account also retains the buffer zone intuition that plays such an important role in Williamson's framework. However, as the last few sections have shown, one can accommodate and explain the tension between inexactness and introspection without the need to reject KK.

[29]See Dokic and Égré (2009:19) for a response to the *ad-hoc* accusation.

## 5 Concluding remarks

Inexact observations are important when it comes to introspection; they are important for dynamic introspection. In sections 2 and 3 I have shown how inexact observations are in conflict with dynamic introspection. In section 4 I have further argued that Williamson reaches a wrong conclusion by suppressing the dynamic aspects of inexact observations. My argument extends to any account that takes epistemic indistinguishability to be non-transitive in situations of inexact observations,[30] and more broadly to any account that, following Williamson, assumes that inexact observations create a synchronic conflict between first- and higher-order evidence. Within his static formulation, Williamson motivates **P2** with the idea that knowledge requires safety, and safety requires an error free buffer zone; knowledge can never be obtained at a world 'on the edge.' My account of inexact observations fully adheres to the buffer zone intuition that motivates Williamson. Every inexact observation event leads to an updated model in which the actual world is surrounded by close but epistemically possible worlds that act as a buffer zone (consult the position of the actual world in Figures 3 to 5). No observation event can put the actual world at the 'edge' of the updated model – and thereby the agent in an epistemically dangerous place. But my account also shows that one can follow the safe buffer zone intuition without accepting a strong premise like **P2**. If one wants to stick with the truth of **P2**, one cannot just cite margin-for-error type safety considerations. Even in situations where the agent is able to use their margin-for-error knowledge to conclude something stronger about what they know, such reasoning cannot be iterated. This is because the margin-for-error principle only describes the epistemic effect of inexact observations; it does not describe the effects of inexact observations after they are combined with non-observational knowledge. In such cases, knowledge of margin-for-error cannot be used to come to know anything stronger.

A more general conclusion is that in the context of an externalist epistemology, reasoning about epistemic updates requires special care. In standard dynamic epistemic logic, update operators are *transparent* to the agent, in the sense that the behavior of the update is the same inside and outside the scope of knowledge.[31] This idealized assumption is exemplified in two DEL axioms, known as *no-miracles*, $\langle\varphi\rangle K\psi \to K[\varphi]\psi$ and *perfect recall*, $K[\varphi]\psi \to [\varphi]K\psi$ (van-Benthem et al. 2009,

[30] E.g. Elga 2013, Salow and Ahmed 2017.

[31] A similar point holds for Bayesian epistemology, in which it is assumed that the posterior epistemic state is transparent to the agent prior to the update (as a prior conditional state).

Cao and Wang 2013). Note how the two axioms allow to switch the order of the knowledge and update modalities. In my account of *inexact updates*, these two axioms fail. This is important, as it suggests that within externalism, reasoning about epistemic updates is not transparent, but *opaque* (Cohen 2020b). This should make sense: updates behave differently at the good and bad case, and when we don't know whether we are in the good case, there is no reason to assume that the update will behave in the ways we expect. We don't always know how new evidence will affect us, nor do we always know what evidence brought us to our current state. This *opacity* is crucial for understanding externalist theories of knowledge and some of the puzzles associated with them.

**Acknowledgments:** Versions of this paper were presented in the 2019 Glasgow graduate conference in epistemology and mind, the 2019 Tübingen Masterclass in Theoretical Philosophy with Timothy Williamson, and the Knowledge and its Limits at 20 conference in the University of Geneva. I thank the audiences and commentators for helpful discussion. In addition, I would like to thank Johan van Benthem, Ray Briggs, and Krista Lawlor for the guidance and the many comments they offered in the process of writing this paper. I would also like to thank the anonymous referees who provided detailed comments on the current and earlier versions of this paper.

## Appendix: Deriving an absurdity in the unmarked clock example with dynamic introspection

Recall that $e$ abbreviates $p_{16} \vee p_{17} \vee p_{18}$ and nothing else. We assume the principles:

| | |
|---|---|
| $[e]Ke$ | **main assumption** |
| $K[e](K\neg p_{15} \rightarrow \neg p_{16})$ | **P2′** |
| $[\varphi]K\varphi \rightarrow K[\varphi]K\varphi$ | dynamic introspection |
| $\langle\varphi\rangle\psi \leftrightarrow [\varphi]\psi \wedge \varphi$ | **Partial function** |

We further assume that both $K$ and $[\varphi]$ obey:

| | |
|---|---|
| $\vdash \varphi \Rightarrow \vdash \Box\varphi$ | nec. rule |
| $\Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$ | $\rightarrow$-distribution |
| $\Box(\varphi \wedge \psi) \leftrightarrow (\Box\varphi \wedge \Box\psi)$ | $\wedge$-distribution |

First we show that update-knowledge closure holds:

**Lemma 1.1:** $\vdash K[\varphi]K(\alpha \to \beta) \wedge K[\varphi]K\alpha \to K[\varphi]K\beta$

*Proof.*

| | |
|---|---|
| $\vdash K(\alpha \to \beta) \wedge K\alpha \to K\beta$ | theorem of epistemic logic |
| $\vdash [\varphi](K(\alpha \to \beta) \wedge K\alpha \to K\beta)$ | nec. rule of $[\varphi]$ |
| $\vdash [\varphi](K(\alpha \to \beta) \wedge K\alpha) \to [\varphi]K\beta$ | distribution of $[\varphi]$ |
| $\vdash [\varphi]K(\alpha \to \beta) \wedge [\varphi]K\alpha \to [\varphi]K\beta$ | distribution of $[\varphi]$ |
| $\vdash K([\varphi]K(\alpha \to \beta) \wedge [\varphi]K\alpha \to [\varphi]K\beta)$ | nec. rule of $K$ |
| $\vdash K[\varphi]K(\alpha \to \beta) \wedge K[\varphi]K\alpha \to K[\varphi]K\beta)$ | distribution of $K$ |

□

Now, we show how to derive the problematic conclusion $K[e]\neg p_{16}$ (in line 10) from our assumptions. Line 6. establishes what I called earlier the auxiliary assumption.:

| | |
|---|---|
| 1. $[e]Ke$ | **main assumption** |
| 2. $K[e]Ke$ | by (DI) |
| 3. $e \to \neg p_{15}$ | analytic truth (or model validity) |
| 4. $K(e \to \neg p_{15})$ | nec. rule of $K$ |
| 5. $[e]K(e \to \neg p_{15})$ | nec. rule of $[\varphi]$ |
| 6. $K[e]K(e \to \neg p_{15})$ | nec. rule of $K$ |
| 7. $K[e]K\neg p_{15}$ | by Lemma 1.1 |
| 8. $K[e](K\neg p_{15} \to \neg p_{16})$ | assumption |
| 9. $K([e]K\neg p_{15} \to [e]\neg p_{16})$ | $\to$-distribution for [ ] |
| 10. $K[e]\neg p_{16}$ | $\to$-distribution for K |

## References


– Baltag, Alexandru and van-Benthem, Johan. (2018). Some thoughts on the logic of imprecise observations. Manuscript.

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

– Blackburn, Patrick , de Rijke, Maarten & Venema, Yde (2002). *Modal Logic*. Cambridge University Press.

– van-Benthem, Johan (2003). Conditional probability meets update logic. *Journal of Logic, Language and Information*. 12 (4):409-421.

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

– Bonnay, Denis and Egre, Paul (2008). Margins for Error in Context. in M. Garcia-Carpintero & M. Kölbel (eds.), *Relative Truth*, pp. 103-127, Oxford University Press.
– Bonnay, Denis and Egre, Paul (2009). Inexact Knowledge with Introspection. *Journal of Philosophical Logic*. 38 (2), pp. 179-228.
– Bonnay, Denis and Egre, Paul (2011). Knowing one's limits - An analysis in Centered Dynamic Epistemic Logic. in P. Girard, M. Marion and O. Roy (eds), *Dynamic Formal Epistemology*, Springer, pp. 103-126
– Briggs, Ray (2009). Distorted reflection. *Philosophical Review* 118 (1):59-85.
– Brueckner, A., and Fiocco, M.O. (2002. Williamson's Anti-Luminosity Argument. *Philosophical Studies* 110: 285-293.
– Christensen, David (2010). Higher-Order Evidence. *Philosophy and Phenomenological Research* 81 (1):185-215.
– Cohen, Stewart (2003). Basic Knowledge and the Problem of Easy Knowledge. *Phil osophy and Phenomenological Research*, 65(2): 309–329.
– Cohen, Michael. (2020a). The problem of perception and the no-miracles principle. *Synthese*. https://doi.org/10.1007/s11229-020-02772-3.
– Cohen, Michael (2020b). Opaque Updates. *Journal of Philosophical Logic*. https://doi.org/10.1007/s10992-020-09571-8
– Conee, E. (2005). The comforts of home. *Philosophy and Phenomenological Research* 70: 444-451.
– Das, Nilanjan & Salow, Bernhard (2018). Transparency and the KK Principle. *Noûs* 52 (1):3-23.
– Dokic, J., Égré, P.(2009) Margin for error and the transparency of knowledge. *Synthese*. 166, 1–20 .
– Dorst, Kevin (2020). Evidence: a guide for the uncertain. *Philosophy and Phen - omenological Research.*
– Dorst, Kevin (2019). Abominable KK Failures. *Mind*.
– Dutant, Julien (2007). Inexact Knowledge, Margin for Error and Positive Introspection. *Proceedings of Tark XI*.
– Elga, Adam (2013). The puzzle of the unmarked clock and the new rational reflection principle. *Philosophical Studies* 164 (1):127-139.
– Goodman, J. and Salow, B. (2018). Taking a chance on kk. *Philosophical Studies*, 175(1):183–196.
– Greco, Daniel (2014a). Could KK Be OK? *Journal of Philosophy* 111 (4):169-197.
– Greco, Daniel (2014b). Iteration and Fragmentation. *Philosophy and Phenomenological Research*. 88 (1):656-673.
– Greco, D. (2015a). Iteration principles in epistemology i: Arguments for. *Philosophy Compass*, 10(11):754– 764.
– Greco, Daniel (2015b). Iteration Principles in Epistemology II: Arguments Against. Philosophy Compass 10 (11):765-771.
– Greco, Daniel (2016). Safety, Explanation, Iteration. *Philosophical Issues* 26 (1):187-

208.
– Halpern, Joseph Y. (2008). Intransitivity and vagueness. *Review of Symbolic Logic*. 1 (4):530-547.
– Hawthorne, John & Magidor, Ofra (2009). Assertion, Context, and Epistemic Access -ibility. *Mind*. 118 (470):377-397.
– Holguín, Ben (2019). Indicative Conditionals and Iterative Epistemology. Manuscript.
– Holliday, Wesley (2018). Knowledge, Time, and Paradox: Introducing Sequential Epistemic Logic. In *Jaakko Hintikka on Knowledge and Game-Theoretical Semantics*. Springer Verlag.
– Horowitz, Sophie (2014). Epistemic Akrasia. *Noûs* 48 (4):718-744.
– Icard, T. Holliday W. Indicative Conditionals and Dynamics Epistemic Logic. (2017). *Proceedings of the Sixteenth Conference on Theoretical Aspects of Rationality and Knowledge (TARK XVI).*
– Lasonen-Aarnio, Maria (2014). Higher-order evidence and the limits of defeat. *Philoso -phy and Phenomenological Research*. 8(2):314–345.
– Lasonen-Aarnio, Maria (2015). New Rational Reflection and Internalism about Rational -ity. In Tamar Szabo Gendler and John Hawthorne, eds., *Oxford Studies in Epistemol -ogy*, volume 5, 145–171. Oxford University Press.
– Lederman, Harvey (2017). Uncommon Knowledge. *Mind*:fzw072.
– Mott, Peter (1998). Margins for error and the sorites paradox. *Philosophical Quarterly* 48 (193):494-504.
– Liu. 2020. (Un)knowability and Knowledge Iteration. *Analysis.*
– Neta, Ram and Rohrbaugh, Guy (2004). Luminosity and the safety of knowledge. *Pacific Philosophical Quarterly* 85 (4):396–406.
– Okasha, S. (2013). On a flawed argument against the KK principle. *Analysis* 73 (1):80-86.
– Roush, Sherrilyn (2017) Epistemic Self-Doubt . *Stanford Encyclopedia of Philosophy*.
– Salow, Bernhard & Ahmed, Arif (2017). Don't Look Now. *British Journal for the Phi -losophy of Science*.
– Sharon, Assaf and Spectre, Levi (2008). Mr. Magoo's mistake. *Philosophical Studies* 139 (2):289-306.
– Stalnaker, Robert (2015). Luminosity and the KK thesis. in Goldberg, Sanford C. (ed.) *Externalism, Self-Knowledge, and Skepticism: New Essays*. Cambridge University Press.
– Vogel, Jonathan (2000). Reliabilism Leveled *The Journal of Philosophy*. 97(1): 602–623.
– Wang, Yanjing & Cao, Qinxiang (2013). On axiomatizations of public announcement logic. *Synthese* 190 (S1).
– Weisberg, Jonathan (2007). Conditionalization, Reflection, and Self-Knowledge. *Philosophical Studies* 135 (2):179-197.
– Williamson, Timothy (2000). *Knowledge and its Limits*. Oxford University Press.
– Williamson, Timothy (2008). Why epistemology cannot be operationalized. In Quentin

Smith (ed.),*Epistemology: New Essays*. Oxford University Press.

– Williamson, Timothy (2013). Gettier cases in epistemic logic. *Inquiry : An Interdisci -plinary Journal of Philosophy* 56 (1):1-14.

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

– Williamson, Timothy (forthcoming). Evidence of Evidence in Epistemic Logic . In in Mattias Skipper Rasmussen and Asbjørn Steglich-Petersen (eds.), *Higher-Order Evi -dence: New Essays*, Oxford: Oxford University Press.

– Yli-Vakkuri, J. and Hawthorne, J. (forthcoming). Being in a Position to Know. *Philosophical Studies.*