Knowing one's limits

An analysis in Centered Dynamic Epistemic Logic

Denis Bonnay & Paul Égré

Abstract By reflecting on the limits of one's knowledge, it may be possible to acquire further knowledge that was unavailable prior to reflection. We propose a dynamic analysis of the process whereby a subject reflects on the reliability conditions of her perceptual knowledge. Following earlier work on Centered Semantics for epistemic logic, we show how to combine it with Dynamic Epistemic Logic and introduce a variation on the standard semantics for epistemic updates. The resulting logic, Centered Dynamic Epistemic Logic, is then applied to analyze the Margin for Error paradox due to T. Williamson. We argue that the paradox stems from a disputable assumption about the properties of consecutive estimates of one's margins for error. We provide a precise characterization of the demarcation between paradoxical and non-paradoxical scenarios in terms of the properties of sequences of estimates.

1 Dynamic Logic and Epistemic Paradoxes

Dynamic epistemic logic has been used to explain away various epistemic paradoxes. Van Benthem [1] showed how the difference between successful and unsuccessful epistemic updates can account for the Fitch paradox. Gillies [8] proposed a similar approach to Moore's paradox, and Gerbrandy [7] recently examined the Surprise Examination paradox in the light of dynamic epistemic logic. In all three cases, the paradoxes can be seen to originate in an equivocation between what one may learn or *realize*, and what one may *actually* know. In the most typical case, the

---

Denis Bonnay
University Paris-Ouest, IREPH/IHPST - Département d'Etudes Cognitives de l'ENS, 29 rue d'Ulm, 75005 Paris, France. e-mail: denis.bonnay@ens.fr

Paul Égré
Institut Jean-Nicod (CNRS/ENS/EHESS) - Département d'Etudes Cognitives de l'ENS, 29 rue d'Ulm, 75005 Paris, France. e-mail: paulegre@gmail.com

case of Moore's paradox, the agent is assumed to learn that a certain fact holds, of which she was not aware. The fact makes crucial reference to the very ignorance of the agent, so that realizing it results in the fact holding no more.

As an example, suppose that it is raining and the agent does not know it. The agent is told so. She realizes that it is raining and that she does not know it. But this changes her epistemic state, in such a way that it is no longer true that she does not know that it is raining. Thus, realizing that $p \wedge \neg Kp$ does not yield knowledge of $p \wedge \neg Kp$. The notion of *epistemic update* used in Dynamic Epistemic Logic can be used to account for that situation. In Dynamic Epistemic Logic, whenever an agent is informed about some *atomic* fact (true proposition), she thereby comes to know that it is true. Such updates on the agent's epistemic state are called *successful*, when updating by some proposition leads to the knowledge of that proposition. Not all updates with true propositions need be successful, however, in particular when *non-atomic* propositions are involved. The three paradoxes we mentioned all involve unsuccessful updates in that sense, namely updates by non-atomic true propositions that, when announced or revealed to the agent, are no longer true *after* they have been announced. As argued by van Benthem, the logical analysis of these paradoxes in dynamic logic suggests that their paradoxical flavor primarily stems from the illusion that all updates should be successful.

In [3], we introduced a non-standard semantics for epistemic logic, Centered Semantics, as well as a further generalization called Token Semantics, in order to account for another epistemic paradox, originally due to T. Williamson, and akin to the Surprise Examination paradox. From a model-theoretic point of view, the new semantics was designed to make compatible a notion of inexact knowledge, based upon non-transitive and non-euclidian relations of epistemic accessibility, and the principles of positive and negative introspection, whose validity is equivalent to transitivity and euclideanness of the accessibility relation. At the conceptual level, we intended the semantics to dispel what we saw as an excessive tension between epistemic principles that are plausible when viewed separately, but conflicting when brought together. Thus, Williamson showed that margin for error principles for knowledge (principles of the form: I know that p provided p is true in all contexts sufficiently similar to the actual one), plus knowledge of these principles, are not compatible with positive introspection. Williamson considers this a *reductio* of the introspection principles. Centered semantics, on the other hand, has the remarkable feature that it can validate the principle of margin for error without thereby validating knowledge of the principle.[1] We argued that this was the way to go: 'knowledge of the margins' results in a change of the margins, so that in general, it is not safe to assume that an agent knows that her knowledge obeys a fixed margin. In our perspective, Williamson's paradox should be taken as a *reductio* of the knowability of margin for error principles, rather than as a case against introspection.

However, our formal account of the epistemic scenario underlying the paradox remained *static*. We merely showed how to resist knowledge of the margin for error

[1] This is a particular case of a general failure of the rule of necessitation over models, see [3] for details.

principle while accepting the principle as a valid principle constraining the semantics for knowledge. Nevertheless, the conceptual argument presented in favor of this strategy was essentially *dynamic*. In [3], we point out that reflection on the limitation of one's knowledge makes it possible to improve on that knowledge. Such reflection is obviously a good thing, but it makes dubious the assumption that knowledge about one's limits is unmodified through the reflection process. In [3], however, we did not present a worked out elaboration of this dynamic intuition. In this respect, our proposal to construe Williamson's paradox as a *reductio* of knowledge of the margin for error principle remained incomplete.

Given the structural similarities between Williamson's paradox and the Surprise Examination paradox, which are fully explicit in [12], and given the broader similarities between our conceptual analysis of Williamson's paradox and previous accounts of epistemic paradoxes based on dynamic epistemic logic, it is worth considering whether a complete account of Williamson's paradox can be provided by merging Dynamic Epistemic Logic and Centered Semantics. This is precisely what this paper achieves: we propose a centered dynamic epistemic logic, for which the standard axiomatization remains sound and complete, and we show how it can be used to account for the dynamics of reflection on one's margins.

An important fact about this merge is that we could not have done without centered semantics in the first place: an account based solely upon Dynamic Epistemic Logic and Kripke semantics would not preserve the introspection principles over models of inexact knowledge. Furthermore, the merging creates values for shareholders on both sides. As explained, short of introducing epistemic updates, the reflection process on one's epistemic limitations remained unaccounted for in Centered Semantics. Conversely, the issue of what happens when one reflects upon one's epistemic limitations constitutes an original field of application for Dynamic Epistemic Logic. The Fitch paradox or knowability paradox, which can be seen as derivative from Moore's paradox, looms quite large in discussions on the limits of knowledge. Prima facie, limitations imposed by Moorean sentences on our knowledge can be quarantined. However, failure to distinguish between perceptual alternatives is a very pervasive phenomenon. And so is to some extent the reflection on our limitations: in everyday life, we constantly try to maximize our knowledge by taking into account what we realize that we do not know.

2 Centered Semantics with an update operator

In this section we present a logic that we call Centered Dynamic Epistemic Logic (CDEL for short). The language of CDEL is the same as the language of DEL, namely an epistemic language with a dynamic update operator. The semantics differs from that of DEL in two respects. Regarding the static knowledge operator, the underlying semantics is Centered Semantics (CS) instead of standard Kripke semantics. Because of that, the semantics of the update operator requires some minor adjustments. In what follows, we first present CS for the basic epistemic language.

In the second part, we define CDEL, the dynamic version of CS, for the epistemic language with an update operator. In all that follows, the language $\mathscr{L}$ of static epistemic logic is defined by $\phi := p|\neg\phi|(\phi \wedge \phi)|K\phi$, where K is the static knowledge operator. The language $\mathscr{DL}$ of dynamic epistemic logic is the extension of $\mathscr{L}$ defined by: $\phi := p|\neg\phi|(\phi \wedge \phi)|K\phi|\ [\phi]\phi$, where $[\,]$ is the update operator. The notation $\langle\phi\rangle$, which we shall use below, is an abbreviation for $\neg[\phi]\neg$, namely for the dual of the update operator.[2]

2.1 Centered Semantics

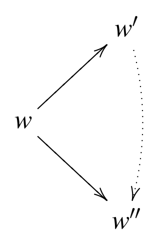
Centered Semantics is a two-dimensional semantics in which epistemic alternatives are relativized to the actual world: when an epistemic operator is evaluated relative to a world that is possible for all we know, the accessible worlds are taken to be those worlds that are already accessible from the actual world. We recall here the definition of truth in CS. It is a two stage definition. Truth is first defined with respect to couples of worlds, and truth at a single world is then defined by diagonalizing. Here are the precise definitions:

Definition 1. Truth for couples of worlds:

(i) $\mathscr{M},(w,w') \models_{\text{CS}} p$ iff $w' \in V(p)$.
(ii) $\mathscr{M},(w,w') \models_{\text{CS}} \neg\phi$ iff $\mathscr{M},(w,w') \not\models_{\text{CS}} \phi$.
(iii) $\mathscr{M},(w,w') \models_{\text{CS}} (\phi \wedge \psi)$ iff $\mathscr{M},(w,w') \models_{\text{CS}} \phi$ and $\mathscr{M},(w,w') \models_{\text{CS}} \psi$.
(iv) $\mathscr{M},(w,w') \models_{\text{CS}} K\phi$ iff for all w'' such that wRw'', $\mathscr{M},(w,w'') \models_{\text{CS}} \phi$.

Definition 2. $\mathscr{M},w \models_{\text{CS}} \phi$ iff $\mathscr{M},(w,w) \models_{\text{CS}} \phi$

Clause (iv) of the definition accounts for the "centered" feature of the semantics. Looking on a picture,



we can easily see how clause iv) works. If we evaluate a formula at a world w, and if the evaluation process takes us to a world w' accessible from w, we then take as worlds w'' accessible from w' the worlds that are in fact accessible from w.

[2] See [4], whose notational conventions are taken up here. More precisely, the language $\mathscr{DL}$ corresponds to the language $\mathscr{L}_{K[]}$ of Public Announcement Logic defined in [4], p.73, for the case of a single agent.

Let us call 'evaluation world' the world at which a whole formula is being evaluated, as opposed to other worlds visited during the evaluation process. Semantic evaluation is *centered* on the evaluation world in the following sense: at every step in the evaluation process, the worlds that are accessible are the worlds accessible from the evaluation world. In particular, clause iv) entails that for every w and w': $\mathscr{M},(w,w') \vDash_{CS} K\phi$ iff $\mathscr{M},(w,w) \vDash_{CS} K\phi$ iff $\mathscr{M},w \vDash_{CS} K\phi$.

K45 is sound and complete with respect to CS on the class of all frames, and **S5** is sound and complete with respect to CS on the class of reflexive frames (see [3]). Centering makes positive and negative introspection automatically satisfied over arbitrary structures. Thus **K45** axioms and theorems turn out to be valid even over non-transitive and non-euclidian frames, just as **S5** axioms and theorems turn out to be valid even on reflexive frames where the accessibility relation is not an equivalence relation.

2.2 Centered Dynamic Epistemic Logic

From a static viewpoint, the rationale for using CS rather than standard Kripke semantics is to be able to preserve the introspection principles over structures of inexact knowledge, where the agent's information is not sharply distributed into equivalence classes of worlds (we refer to [3] for ampler discussion). Taking a dynamic perspective, what we seek now is to provide an analysis of epistemic updates within the framework of CS. The first question on our agenda is of course: how shall we define epistemic updates in this setting?

Let us briefly recall how things work in standard DEL. As explained above, a new dynamic action modality is added to basic epistemic logic. If ϕ and ψ are formulas, so is now $[\phi]\psi$. The subformula $[\phi]$ means *updating* with the information that ϕ, and $[\phi]\psi$ is true if ψ is true after updating with ϕ. For our purposes, in the case of a single agent, $[\phi]$ will mean that the agent *realizes* that ϕ is the case. In particular, $[\phi]K\psi$ states that the agent will know that ψ after realizing ϕ, namely after receiving the hard information that ϕ. $[\phi]$ itself is interpreted by an operation on models. Basically, all worlds which are not ϕ are cut off, and the standard clause reads:

(*) $\mathscr{M},w \vDash [\phi]\psi$ iff, if $\mathscr{M},w \vDash \phi$, then $\mathscr{M}|\phi,w \vDash \psi$

where $\mathscr{M}|\phi$ is the restriction of $\mathscr{M}$ to ϕ worlds. The condition that $\mathscr{M},w \vDash \phi$ guarantees that we are talking about a truthful piece of information.

So what about updates in centered semantics? The natural clause to consider would be:

$$\mathscr{M},(w,w') \vDash_{CS} [\phi]\psi \text{ iff, if } \mathscr{M},(w,w') \vDash_{CS} \phi, \text{ then } \mathscr{M}|\phi,(w,w') \vDash_{CS} \psi$$

But one question arises: what do we mean exactly by $\mathscr{M}|\phi$? The question arises because truth has been relativized in CS, so that various options are available. In this particular context, we can restrict $\mathscr{M}$ either to worlds w'' such that $\mathscr{M},(w'',w'') \vDash \phi$

or to worlds w'' such that $\mathscr{M},(w,w'') \vDash \phi$. The second option is the more natural: accessibility remains relativized to the evaluation world w. The adequacy of this option will show up in subsequent theorems about updates in centered semantics. By contrast, choosing the first option would permit "cheating" through epistemic updates, namely worlds that are not direct alternatives to the actual world could become relevant during the evaluation.

To make this clear, we shall therefore write the clause for epistemic updates as clause (v):

(v) $\mathscr{M},(w,w') \vDash_{\mathrm{CS}} [\phi]\psi$ iff, if $\mathscr{M},w \vDash_{\mathrm{CS}} \phi$, then $\mathscr{M}|\phi_w,(w,w') \vDash_{\mathrm{CS}} \psi$

where $\mathscr{M}|\phi_w$ is $\mathscr{M}$ restricted to worlds w'' such that $\mathscr{M},(w,w'') \vDash_{\mathrm{CS}} \phi$.

By definition, the semantics DEL for $\mathscr{DL}$ is the basic Kripke semantics augmented with clause (*) for the update operator. We define CDEL for $\mathscr{DL}$ as the basic Centered Semantics (clauses (i)-(iv)) augmented with clause (v) for the update operator. When no assumptions are made on the epistemic accessibility relation, DEL is axiomatized by the logic **K** plus the following recursion axioms *RA* (see [4]):

$$
\begin{array}{ll}
[\phi]p & \leftrightarrow \phi \rightarrow p \\
[\phi]\neg\psi & \leftrightarrow \phi \rightarrow \neg[\phi]\psi \\
[\phi](\psi \wedge \chi) & \leftrightarrow [\phi]\psi \wedge [\phi]\chi \\
[\phi]K\psi & \leftrightarrow \phi \rightarrow K(\phi \rightarrow [\phi]\psi)
\end{array}
$$

In the static case, we saw that CS is axiomatized by **K45** over the class of all frames. Likewise, in the dynamic case, CDEL is axiomatized by **K45** and the recursion axioms characteristic of DEL over the class of all frames:[3]

Theorem 1 **K45** *(resp.* **S5***) plus the recursion axioms is sound and complete with respect to Centered Semantics with updates on the class of all frames (resp. of all reflexive frames).*

What the theorem shows is that CDEL, just like CS, is conservative in terms of axioms, though innovative in terms of models. This is a desirable feature, since our aim in the next section is to use the centered version of dynamic logic to analyze Williamson's paradox in a way that still combines introspection and a notion of inexact knowledge, but taking updates into account.

3 The Margin of Error paradox

What we call a paradox is not intended as such by Williamson, but rather as an argument against the principle of positive introspection, and as part of a more general philosophical argument against the so-called *luminosity* of mental states, namely the

[3] The proof is given in the Appendix.

idea that there might be propositions ϕ that are automatically known, namely such that $\phi \rightarrow K\phi$ (thus when ϕ is of the form $K\psi$, positive introspection becomes a particular instance of luminosity). Yet what Williamson's argument establishes is that a number of independently plausible premises lead to contradiction. The form of the argument is also closely related to that of a sorites argument. Because of these two features, the argument in its general form has sometimes been called the *luminosity* paradox (see [9], [6]). To give it a name here, we shall call it the *Margin of Error* paradox, because the notion of a margin of error used by Williamson is distinctive of its formulation.

3.1 The paradox

The form of the argument is the following: Mr Magoo is a myopic character who observes a tree at some distance. Magoo is certain that the tree is less than k meters high (say less than $k = 20$). Magoo's knowledge about sizes is constrained by a *margin for error principle*, whereby for Magoo to know that the tree is less than k meters high, the tree has to be less than $k-\eta$ meters high, for a particular η. For instance, suppose $\eta = 1$: the principle says that for Magoo to know that the tree is less than 20 meters high, the tree cannot measure 19 meters, for from where he is Magoo cannot reliably discriminate between sizes that differ by just 1 meter. By assumption, Magoo is also taken to be *aware of this limitation* on his knowledge, to be *positively introspective*, and to *know the consequences* of what he knows. For the argument to go, finally, η must be positive, but can be arbitrarily small.

The structure of the argument is the following. Let '$(s < k)$' be an atomic proposition standing for 'The tree is less than k meters high'. Similarly, '$(s \geq k)$' is an atomic proposition standing for 'The tree is at least k meters high'. Starting from assumption (1), (4) follows via (2) and (3) on the basis of the assumed principles for Magoo's knowledge:

(1) $K(s < k)$
(2) $K(s \geq (k-\eta) \rightarrow \neg K(s < k))$
(3) $KK(s < \eta)$
(4) $K(s < (k-\eta))$

(1) is the assumption that Mr Magoo knows the tree to be less that k meters high. If the conditions of Mr Magoo's woodsy observation are sufficiently good and if k is taken to be sufficiently large, (1) is quite uncontroversial. (2) expresses Mr Magoo's knowledge about his own margin for error, namely that it is at least η. (3) follows from (1) by positive introspection. (4) follows from (2) and (3) by closure of knowledge under logical consequence. Here η can be taken to be arbitrarily small. By repeating the argument i times, one may reach the conclusion that Mr Magoo knows the size of the tree to be less than $k - i \cdot \eta$. So long as (2) is taken to hold without restriction for all relevant values of k, one can repeat the argument enough

times to reach an absurd conclusion with respect to the actual size of the tree (namely that the tree is of size 0).

The principle of closure under logical consequence is not objectionable in this context, at least as an idealization holding for rational agents – who might have to evaluate tree sizes just as forest wardens do. So there are only two options left: either the principle of positive introspection is to be rejected, thereby blocking the inference from (1) to (3), or (2) is to be rejected, suggesting that one cannot improve on (1) by reflecting on one's limitations. Stressing the fact that η – the estimated margin – can be taken to be arbitrarily small, Williamson construes the paradox as a *reductio* of positive introspection. By contrast, the gist of our dynamic analysis will be to point (2) as the real culprit for the paradox.

The problem is that rejecting (2) seems counterintuitive at first glance. To begin with, the principle of margin for error itself seems quite reasonable. I certainly cannot know by vision alone that a tree that is $k-\eta$ meters tall is less than k meters tall when a difference of η meters in size is too small to be detected by my eye. And it is certainly possible to find a value for η – say 0.01 meters – such that no such difference meets the eye. More generally, our perceptual knowledge is certainly bounded by the limitations of our perception, namely by what we cannot perceptually discriminate.

If the principle of margin for error holds, it seems equally reasonable to grant that we can become aware of this fact. I can certainly realize that my eyesight is far from perfect. Reflecting on this limitation, I can realize that my visual knowledge obeys a margin for error principle. I know for sure that if a tree is 11.99 meters tall, I cannot know for sure that it is less than 12 meters tall. Hence, if (I know that) I know that the tree is less than 12 meters tall, I thereby know that it is also less than 11.99 inches tall, and so starts the sorites.

3.2 Knowing and realizing

The paradoxicality of Williamson's argument originates from the fact that the reasoning seems perfectly valid and the premises sound. Or...could we have been misled? In [5], Dokic & Égré argue that margins for error come in different varieties. My initial knowledge that the tree is less than 12 meters tall is purely visual. My acquired knowledge that it is less than 11.99 meters tall is not so. If this knowledge is to obey the same margin for error principle as my initial visual knowledge, and if I can know it does, we are in trouble. But why should the knowledge I gained, which is based partly on perception, partly on rational reflection and on drawing inferences, be subject to exactly the same limitations as my initial knowledge? On the contrary, it seems that whatever limitations my visual knowledge was subject to, my reflective knowledge is not subject to *exactly those*, since my reflective knowledge *improves on* my visual knowledge.

The strategy in [5] was to carefully distinguish between kinds of knowledge according to their sources. Indexing knowledge operators accordingly blocks iterations of the reasoning after the first step. One problem with this strategy, however, is that it sprays 'plain knowledge' into several varieties of knowledge, and it remains silent regarding the principles governing the generic notion of knowledge, irrespective of its source. Indeed, should this general knowledge also obey a margin for error principle, then a revenge form of the paradox would be lurking around.

We wish to preserve the intuition put forward in [5] that the paradox can be explained by observing that there is a somewhat hidden but crucial assumption that the estimated margin can be kept fixed. However, we also want to make this intuition compatible with the original analysis in terms of a single notion of general knowledge involved throughout the argument. The story we want to tell is essentially dynamic. There is some visual knowledge to begin with. Reflection comes in and results in improved (mixed) knowledge. So there is knowledge at the beginning, and there is knowledge at the end. Yet does the transition itself, namely reflection upon one's limits, qualify as a piece of knowledge? Williamson thinks it does, and thinks that it is safe to assume that there is a fixed margin η such that we can always assume that Mr Magoo's visual margin for error can be known by himself to be at least η. By 'always', we mean that (2) is considered to remain true even as Mr Magoo's knowledge of the size of the tree has improved after, say, the first round of reasoning and reflecting on his margin for error.

An alternative analysis treats Mr Magoo's reflection on his limitations as *realizing* something, rather than *knowing* it. At this point, the analogy with Moorean scenarios is telling. *Realizing* that it rains and that I do not know it does not count as *knowing* that it rains and that I don't know it. Indeed, it cannot count as knowledge. The mere fact that one realizes that one does not know something which is true changes the relevant epistemic facts. The same holds of margins of error. Mr Magoo's realizing that his knowledge concerning the tree size is limited changes the relevant epistemic facts. Indeed, he is able to gain new information on the basis of his reflection on his limitations.

3.3 Reanalyzing the paradox with epistemic updates

Following these intuitions, the best tool to model Mr Magoo's scenario is epistemic updates, which model from a dynamic perspective what happens when an epistemic agent realizes that something is the case. Using updates is more adequate than using the K operator of epistemic logic, which models from a static perspective what happens when an agent knows that something is the case.

Let us abbreviate '$(s \geq (k-\eta)) \rightarrow \neg K(s<k)$' by $ME(k,\eta)$. $ME(k,\eta)$ states that Mr Magoo's margin is of at least η when it comes to estimating sizes around k. $\mathscr{M}, w \vDash_{\mathrm{CS}} [\phi]\psi$ does not exactly say that ψ holds in the model one gets from $\mathscr{M}, w$ by updating with ϕ, because $[\phi]\psi$ is trivially true when ϕ is false at w.[4] It says that ψ will be true *if the update is successful*. To express that ψ will hold after

[4] The same is true of course for $\mathscr{M}, w \vDash [\phi]\psi$.

the successful update by ϕ, we need to use the dual of the update operator, namely $\langle\phi\rangle\psi$. For $\mathscr{M},w \models \langle\phi\rangle\psi$ provided $\mathscr{M},w \models \phi$ and $\mathscr{M}|\phi,w \models \psi$.[5] Here then is the new formalization we suggest for the argument:

(1') $K(s<k)$
(2') $ME(k,\eta)$
(3') $K(s<k) \rightarrow [ME(k,\eta)]K(s<(k-\eta))$
(4') $\langle ME(k,\eta)\rangle K(s<(k-\eta))$

On this account, Mr Magoo starts with some knowledge about the size of the tree being less than k (1'). It is a fact that his margin for error is at least η when it comes to estimating heights around k (2'). If Mr Magoo knows something to be of size at least k, and he realizes that his margin for error is at least η when it comes to estimating heights around k, then he will come to know that the size of the tree has to be less than $k-\eta$. Hence after reflecting on his margin, Mr Magoo does know the size of the tree to be less than $k-\eta$ (4').

We shall now briefly compare with the earlier formalization by Williamson. (1') is the same as (1), the first premise in Williamson's argument. (2) was knowledge of a basic margin for error principle, namely that Mr Magoo's margin is of at least η. (2') states the basic margin of error principle, the epistemic use of which is deferred to (3'). (3') describes the reflection process itself and says that it results in a gain of η in terms of knowledge of heights. Getting (4') as a conclusion means that (4) is true in the situation we get *after* an epistemic update on $ME(k,\eta)$ starting from a situation in which (1) is true. This is a significant difference with the previous static account according to which (4) is true in the same epistemic situation as the one in which premise (1) is taken to be true. Note that (4') admits of a natural temporal reading.[6] It says that when epistemic facts are changed according to what it means to realize that the margin was at least η, *then* it becomes known that the size of the tree is less than $k-\eta$.

What about the soundness of the argument?[7] The argument is certainly valid: (4') follows from (1'), (2') and (3') by propositional logic alone. Just as before, (1) can certainly be assumed to be true in some situation, and (2) will be true as well if we agree that knowledge, or at least perceptual knowledge, obeys a margin for error principle. As a consequence, the question whether the argument is sound boils down to the question whether (3'), that is $K(s<k)) \rightarrow [ME(k,\eta)]K(s<(k-\eta))$ is true (in general or in the particular structures modeling the scenarios under scrutiny).

$ME(k,\eta)$ is equivalent to $K(s<k) \rightarrow s<(k-\eta)$, so (3') is of the form $K \rightarrow [Kp \rightarrow q]Kq$. The formula says that if I know p, and if I realize that knowing p implies q, then I know q as well. Should this be fine? Realizing that knowing p implies q will help only if I can use my knowledge that p, that is only if I know that I know p. And then to conclude that q, I need to apply *modus ponens*. So the

[5] See [4], prop. 4.14 p. 78: the formula $\langle\phi\rangle\psi$ is thereby equivalent to $\phi\wedge[\phi]\psi$ and to $\phi\wedge\langle\phi\rangle\psi$.
[6] See [2] on merging epistemic updates and temporal logic.
[7] By a *sound* argument, we mean a valid argument whose premises are true.

general principle of which (3') is an instance seems to be acceptable under two closure assumptions about knowledge, namely that knowledge is closed under logical consequence and that knowledge is introspective. This is no surprise since these two principles were used in Williamson's derivation of the paradox but did not appear explicitly in (1')-(4').

Note that the principle is stated for atoms only. It is crucial that it does not apply to any arbitrary formula. Our informal discussion took for granted that the truth of q is preserved under realizing $Kp \rightarrow q$. This is fine because q describes a non-epistemic fact. But if q were replaced with a formula ψ describing an epistemic fact, realizing that ψ is implied by Kp might result in ψ ceasing to be true. As a consequence, it would be absurd to claim that $K\phi \rightarrow [K\phi \rightarrow \psi]K\psi$ is valid in general. To see this at a glance, take any tautology $\top$ for ϕ, we get $K\top \rightarrow [K\top \rightarrow \psi]K\psi$ which is equivalent to $[\psi]K\psi$. What we get is thus a success principle for ψ: realizing that ψ results in knowing that ψ. As we made clear in the first section, success principles cannot be assumed to be valid no matter what. Take for ψ the Moorean sentence $p \wedge \neg Kp$. $[p \wedge \neg Kp]K(p \wedge \neg Kp)$ is false whenever $p \wedge \neg Kp$ is true to start with, because $K(p \wedge \neg Kp)$ is contradictory.

With this restriction clearly in mind, we think that $Kp \rightarrow [Kp \rightarrow q]Kq$ is on the face of it a quite reasonable assumption, at least for idealized rational agents. It says that I can actually improve on my knowledge by realizing that some inferential connections hold between my knowing of certain things, this is Kp, and some other things, namely q. Denying such a principle would certainly deprive epistemic updates of some of their interest, since it would severely limit our ability to gain knowledge through updates. Eventually, one's attitude towards this principle might depend on its relationships with other principles, and the fact that positive introspection and closure under logical consequence came up in our informal discussion is bound to suggest various pros and cons. We shall not propose here an independent defense of positive introspection and closure under logical consequence. In some contexts, it might indeed be more realistic to assume that they do not hold. But it seems to us that it should be nevertheless always coherent to assume that they both hold. It would be quite surprising if it were the case that the very idea of idealized rational agents, who are able to know what they know and draw all the consequences of what they know, was intrinsically incoherent. Therefore, we favor *a priori* an analysis of the Margin for Error paradox which would show that the paradox can be explained and these principles maintained. This is exactly what we shall offer in the last section of this paper.

Finally, let us state precisely the connection between our schema and the principles used in Williamson's formalization. Given two schematic formulas A and B in the language of Dynamic Epistemic Logic, we say that *B follows from A modulo closure under logical consequence and the recursion axioms for updates* (notation: $A \vdash_{Cl,RA} B$) if and only if we can get any instance of B from instances of A using only propositional logic, closure of knowledge under logical consequence and the recursion axioms for updates. The following fact holds:[8]

---

[8] See the proof in part B of the Appendix. One might have hoped to get as well the converse $Kp \rightarrow [Kp \rightarrow q]Kq \vdash_{Cl,RA} Kp \rightarrow KKp$. As pointed out to us by Olivier Roy, this is however not

Fact 2 $Kp \rightarrow KKp \vdash_{Cl,RA} Kp \rightarrow [Kp \rightarrow q]Kq$

Thus, our formal rendering in terms of epistemic updates might seem to confirm Williamson's idea that positive introspection is to blame for the paradox. The only option, if we want to resist (4'), is to deny (3'), and denying (3') logically implies denying introspection (if closure under logical consequence is granted, and we agree with Williamson that it should be granted). However, we do not have to deny (4'): in most cases, it might indeed be perfectly reasonable to accept (4'). The problem shows up when the argument is iterated. Our take is that it should be at least coherent to assume that the agent knows what he knows, and we wish to explain *why* iterating the argument is problematic in this context. We shall see in the next section that this is precisely the point where using epistemic updates makes a difference.

3.4 CDEL does it better

Up to now, we have only discussed the validity of the argument, in its two different forms, and the intuitive plausibility of the premises. Before discussing iterations of the argument, it is well worth looking at the exact truth conditions of the sentences involved. The point we want to make is that our formalization using epistemic updates makes sense only if the semantics for update operators is given by Centered Semantics instead of the standard semantics. DEL does not make the right predictions on the intended models, and this will be our reason for using later on CDEL rather than DEL in order to provide a model-theoretic analysis of what is going when the argument is iterated.

Let us consider a Kripke model $\mathscr{M}_d = \langle W, R_d, V \rangle$, where W is a space of worlds, each of which is indexed by a real number in $\mathbb{R}^+$. Each world w_r is to be thought of as a world at which Mr Magoo's tree is r meters tall. The valuation V is defined accordingly by letting $w_r \in (s \leq k)$ iff $r \leq k$. Mr Magoo's eyesight is characterized by his ability to tell the difference between any two objects whose sizes differ by at least d meters, where d is a real number greater than zero. d is the margin for error which determines Mr Magoo's visual knowledge, and it can be used to define the accessibility relation R encoding Mr Magoo's knowledge by setting $w_r R_d w_{r'}$ iff $|r - r'| \leq d$. R is symmetric and reflexive, but it is not transitive. Following Williamson in [11], we shall call models like $\mathscr{M}_d$ *margin models*.[9] Margin models are certainly the natural models to use if one thinks of knowledge as being determined by a margin for error principle, so we take them to be the intended models for Mr Magoo's scenarios.

---

the case since one can find a Kripke-structure validating (every instance of) $Kp \rightarrow [Kp \rightarrow q]Kq$ but invalidating $Kp \rightarrow KKp$. We conjecture that it is possible to get a full equivalence by suitably liberalizing the schema $Kp \rightarrow [Kp \rightarrow q]Kq$ so as to allow not only atoms but also any successful formula.

[9] In [11], Williamson considers an arbitrary space of worlds equipped with a metric, and rephrases the semantics so as to appeal directly to the parameter d with no detour through a defined accessibility relation.

Centered semantics and Kripke semantics make different predictions when margin models are used.

Fact 3 *The following propositions hold:*

(i) $\mathscr{M}_d \vDash_{\mathrm{CS}} ME(k,\eta)$*, for all* $k \in \mathbb{R}^+$ *and* $0 < \eta \leq d$.
(ii) $\mathscr{M}_d \nvDash_{\mathrm{CS}} ME(k,\eta)$*, for all* $k \in \mathbb{R}^+$ *and* $\eta > d$.
(iii) $\mathscr{M}_d \vDash_{\mathrm{CS}} K(s<k) \rightarrow [ME(k,\eta)]K(s<(k-\eta))$*, for all* $k,\eta \in \mathbb{R}^+$.
(iv) $\mathscr{M}_d \vDash ME(k,\eta)$*, for all* $k \in \mathbb{R}^+$ *and* $0 < \eta \leq d$.
(v) $\mathscr{M}_d \nvDash ME(k,\eta)$*, for all* $k \in \mathbb{R}^+$ *and* $\eta > d$.
(vi) $\mathscr{M}_d \nvDash K(s<k) \rightarrow [ME(k,\eta)]K(s<(k-\eta))$ *for all k and* $\eta \in \mathbb{R}^+$.

Proof. Proofs are left to the reader and we shall just make two remarks. First, regarding the proof of (iii), it is sufficient to analyze the effect of updates like $[ME(k,\eta)]$ on a model $\mathscr{M}_d, w_r$ where $\mathscr{M}_d, w_r \vDash_{\mathrm{CS}} K(s<k)$. What does the set $\{r' \in \mathbb{R}^+ \ / \ (w_r, w_{r'}) \vDash_{\mathrm{CS}} ME(k,\eta)\}$ look like? Since $\mathscr{M}_d, w_r \vDash_{\mathrm{CS}} K(s<k)$, we have $(w_r, w_{r'}) \vDash_{\mathrm{CS}} K(s<k)$ as well. So we are going to take away all the worlds at which $s < (k-\eta)$ is not true. Thus, $\{r' \in \mathbb{R}^+ \ / \ (w_r, w_{r'}) \vDash_{\mathrm{CS}} ME(k,\eta)\} = [0, k-\eta[$, and now $\mathscr{M}_d \backslash [k-\eta, +\infty[, w_r \vDash_{\mathrm{CS}} s < k-\eta$, since all the worlds where the tree is $k-\eta$ tall or taller have been cut off. (vi) follows from Fact 4 below.

$ME(k,\eta)$ is a statement about an approximation of Mr Magoo's margin. It should hold exactly when the approximation is correct, that is whenever η is indeed smaller than the d such that Mr Magoo cannot distinguish between objects whose size differ by no more than d. So (i), (ii) and (iv) and (v) are welcome properties which are shared by Centered Semantics and standard Kripke semantics. (iii) says that premise (3') holds in margin models when evaluated according to Centered Semantics, and (vi) that it fails to hold (quite generally) when evaluated according to standard Kripke Semantics. This is no surprise since, as we said, (3') follows from positive introspection in normal modal logics and the gist of Centered Semantics is to enforce introspection even on non-transitive models such as margin models.

This difference can be traced to a more general contrast between CDEL and DEL. Let us say that a semantics $\vDash_S$ for a language with update operators is *validity-insensitive* if the following hold: let ϕ be a formula which is valid on a model $\mathscr{A}$, let ψ be an arbitrary formula and w be any world in $\mathscr{A}$, then $\mathscr{A}, w \vDash_S [\phi]\psi$ iff $\mathscr{A}, w \vDash_S \psi$. A semantics will be said to be *validity-sensitive* if it is not validity-insensitive.

Fact 4 *The following propositions hold:*

- $\vDash$ *is validity-insensitive*
- $\vDash_{\mathrm{CS}}$ *is validity-sensitive*

Proof. $\vDash$ is validity-insensitive because if ϕ is valid on $\mathscr{A}$, $\mathscr{A}|\phi$ is the same as $\mathscr{A}$. This is not true of $\vDash_{\mathrm{CS}}$ and the fact that $\vDash_{\mathrm{CS}}$ is validity-sensitive can be checked on a suitably chosen margin model. For example, set $d = 2$, take w_{11} as the actual world. We have $\mathscr{M}_d, w_{11} \nvDash_{\mathrm{CS}} K(s < 12.5)$ but $\mathscr{M}_d, w_{11} \vDash_{\mathrm{CS}} [ME(13.5,1)]K(s < 12.5)$.

Fact 4 brings forward a significant difference between epistemic updates in Centered Semantics and epistemic updates in Kripke semantics. An account of Mr Magoo's scenario in Dynamic Epistemic Logic must capture the intuition that Mr Magoo learns something new when he realizes that his perceptual knowledge obeys a given margin for error. But on margin models, any correct approximation from below of the actual perceptual margin is true everywhere in the model. Therefore, *on these models*, standard Kripke semantics makes the counter-intuitive prediction that realizing that one's knowledge is bounded by a certain margin of error has no epistemic consequence at all. By contrast, as shown in the proof, Centered Semantics correctly predicts that the kind of learning ascribed to Mr Magoo in our scenario does occur, even though the margin for error principle is valid on the considered model. In this respect, Centered Semantics gives a more adequate picture of learning than ordinary Kripke semantics does with regard to margin models.[10]

A complete comparison of the merits of each semantics based on its predictions on margin models should include a discussion of yet another difference.

Fact 5 *The following propositions hold:*

(vii) $\mathscr{M}_d \not\models_{CS} KME(k,\eta)$*, for all* $k \in \mathbb{R}^+$ *and* $0 < \eta \leq d$.
(viii) $\mathscr{M}_d \models KME(k,\eta)$*, for all* $k \in \mathbb{R}^+$ *and* $0 < \eta \leq d$.

This has been discussed to some length in [3], and we refer the interested reader to our earlier paper. Note that failure of $KME(k,\eta)$ is not as strange as it might seem at first sight, if $ME(k,\eta)$ is something for us to realize, and not something for us to know, as it happens with Moorean sentences. (vii) says that premise (2) is false.[11] We shall not press this point here, but shall rather argue that (1)-(4) are simply not the best way to formalize what is going on.

4 Keeping on reflecting

4.1 Once versus more than once

In the previous section, we have offered an alternative to Williamson's analysis of Mr Magoo's inferential story. But what is the point of reanalyzing the argument, if

[10] The difference with respect to epistemic updates mirrors a similar difference with respect to the knowledge operator. In Kripke semantics, if $\mathscr{A} \models \phi$ then $\mathscr{A} \models K\phi$ (namely the rule of necessitation is valid over models, or *model-valid*, see [3]). This is not true in Centered Semantics, which is why learning can occur. When assessing the superiority of Centered Semantics, one should nonetheless keep in mind that it is always possible to change the underlying models. What Mr Magoo learns can be described in classical Kripkean terms on a model in which $ME(k,\eta)$ is not true everywhere. The unravelling of $\mathscr{M}_d$, as described in [3], would yield such a model. The fact remains true that the non-transitive model $\mathscr{M}_d$ validating $ME(k,\eta)$ is arguably the most intuitive and simple formal rendering of Mr Magoo's predicament.

[11] In [3], we welcomed failure of KME on margin models as a way to resist Williamson's argument, but we failed to provide a complete alternative logical analysis of Mr Magoo's scenario.

in both cases one reaches basically the same conclusion, namely that Mr Magoo's knowledge improves, under basically the same assumption, namely that introspection holds? As we have suggested, the essential difference shows up when it comes to iterating the argument – recall that there is nothing intrinsically paradoxical with Mr Magoo's reasoning at the first step and that the paradox comes up when the reasoning is repeated an arbitrary number of times.

If we assume (2) to be true in its general form, that is $K\forall x((s \geq (x-\eta) \rightarrow \neg K(s<x))$,[12] the truth of (1) and (2) leads us to the truth of (4), via introspection. (4) can then replace (1) as a premise, (2) gets instantiated with $k-\eta$ instead of k and we can finally derive the truth of $K(s<k-2\eta)$ by introspection again. If the initial argument is sound, every iteration of it is sound. After i iterations, for a large enough i, we reach the paradoxical conclusion that $K(s<k-i\cdot\eta)$. This is to us one of the main reasons to reject the formalization by (1)-(4). Intuitively, it is perfectly fine for Mr Magoo to reflect at least once on his limitations. The problem – which is still in need of a precise characterization – comes up when we somehow assume that Mr Magoo can go on like that forever. Therefore any formalization which has it that *if the argument is sound once it is forever sound*[13] seems to us to be misguided.

What happens if the reasoning is iterated along the lines of (1')-(4')? Let us have a look at the following continuation:

(4') $\langle ME(k,\eta)\rangle K(s<(k-\eta))$ (as before, the conclusion of the first argument is the first premise of the second)
(5') $\langle ME(k,\eta)\rangle ME(k-\eta,\eta)$
(6') $\langle ME(k,\eta)\rangle K(s<(k-\eta)) \rightarrow [ME(k-\eta,\eta)]K(s<(k-2\eta))$
(7') $\langle ME(k,\eta)\rangle\langle ME(k-\eta,\eta)\rangle K(s<(k-2\eta))$

(5') states that the principle of margin for error with parameters $k-\eta$ and η holds in the new epistemic state obtained trough the previous update. (6') is the second step of Mr Magoo's reflexive process. This time $s<(k-\eta)$ is what is initially known and $s<(k-2\eta)$ is the further information that might be inferred by reflection. (7') gives the conclusion that after *two* steps of reflection, Mr Magoo comes to know that $s<(k-2\eta)$.

Is the continuation of the argument sound? Validity is preserved under the scope of a $\langle\ \rangle$ operator. Since (4'), (5') and (6') are exactly analogous to (1'), (2') and (3'), the argument from (4')-(6') to (7') must be valid. But, of course, soundness is not necessarily preserved. (1')-(3') guarantee that (4') is true, since the argument was sound. If introspection holds, (6') will be true as well. But the case of (5'), that is $\langle ME(k,\eta)\rangle ME(k-\eta,\eta)$, is more involved. The epistemic state obtained after successfully updating with $ME(k,\eta)$ is different from the initial epistemic state. The fact that $ME(k,\eta)$ holds in the initial state, even in the generalized form $\forall x\, ME(x,\eta)$, does not ensure that $ME(k-\eta,\eta)$ holds. $\langle\forall x\, ME(x,\eta)\rangle\forall x\, ME(x,\eta)$

[12] $K\forall x((s \geq (x-\eta) \rightarrow \neg K(s<x))$ is to be construed as equivalent to $\bigwedge_{i\in\mathbb{R}^+} K((s \geq (i-\eta) \rightarrow \neg K(s<i))$. It seems that nothing important hinges on how the details of this quantification over possible heights are spelled out.

[13] This notion is further elaborated in the next subsection, under the name of 'iterative soundness'.

means that updating with ME is successful.[14] If it were the case, we would have that *if the argument is sound once it is forever sound*. But this is not so:

Fact 6 *There are margin models* $\mathscr{M}_d$ *and estimates* η *such that* $\mathscr{M}_d \not\models_{CS} \langle \forall x\, ME(x,\eta) \rangle \forall x\, ME(x,\eta)$.

Proof. Updating with $\forall x\, ME(x,\eta)$ shrinks the margin by η. But then if $\eta > d - \eta$, $\forall x\, ME(x,\eta)$ will be false in the updated model.

In order to illustrate this fact, consider a discrete margin model $\mathscr{M}$ such that $W = \mathscr{N}$, and $d = 1$. Let p_i be the proposition true exactly at index i, and for simplicity, let us write $\mathscr{M}, i \models i$ instead of $\mathscr{M}, i \models p_i$. The model satisfies all margin principles of the form $K\neg(i+1) \rightarrow \neg i$, both relative to Kripke semantics and to centered semantics. In particular, $\mathscr{M}, 17 \models_{CS} K\neg 19 \rightarrow \neg 18$. Unlike with Kripke semantics, $\mathscr{M}, 17 \models_{CS} \langle K\neg 19 \rightarrow \neg 18 \rangle K\neg 18$. However, $\mathscr{M}, 17 \not\models_{CS} \langle K\neg 19 \rightarrow \neg 18 \rangle K\neg 18 \rightarrow \neg 17$. The reason is that $\mathscr{M}|(K\neg 19 \rightarrow \neg 18)_{17}$ does not contain the pair $(17,18)$, hence $\mathscr{M}|(K\neg 19 \rightarrow \neg 18)_{17}, 17 \models_{CS} K\neg 18$, but clearly $\mathscr{M}|(K\neg 19 \rightarrow \neg 18)_{17}, 17 \not\models_{CS} \neg 17$. For instance, after realizing that if I know the size of the tree is not 19, it is not 18 either, it is no longer true that if I now know the size not to be of 18, it should not be of 17.

Fact 6 makes it clear that, even though adequate margin principles are valid on margin models, they may become false when the agent realizes that they hold. Again, the intuition, which is fully accounted for in CDEL, is that realizing that the margin is at least η amounts to diminishing the margin, which might end up being less than η. Therefore, in contrast to the formalization in terms of (1) to (4), the suggested formalization of Mr Magoo's reasoning in dynamic logic makes a substantial difference between going just *once* through the reasoning and repeating it a certain number of times. It could happen that a true conclusion is reached from true premises by running the argument for the first time, whereas a false conclusion is reached later on. It would mean that one of the extra-premises needed to get that conclusion is false, and this is compatible with all the initial premises being true.

We regard this diagnosis of the paradox as fairly plausible. Mr Magoo can certainly reflect on his perceptual limitations and thus acquire knowledge. It would make little sense for us to deny that. But when this reflective process is captured in terms of knowledge about one's absolute margin for error – this is premise (2) – we get the unwelcome consequence that recognizing Mr Magoo's one shot reasoning as correct commits us to accepting each further repetition of this process as equally correct. However, if we can accept the truth of the premises of the argument *and* its validity without being committed to arbitrarily many iterations of it, we no longer have to reject any of the general principles making the reasoning valid. This way introspection can be safe from blame.

Moreover, failure of $\langle \forall x\, ME(x,\eta) \rangle \forall x\, ME(x,\eta)$ fits perfectly the main theme in Dokic and Égré [5], namely that the margin for error corresponding to Mr Magoo's

---

[14] See [4] on successful and unsuccessful updates.

perceptual *and* inferential knowledge is simply not the same as the margin for error corresponding to the purely perceptual knowledge which is Mr Magoo's initial endowment.[15] As a consequence, a correct approximation from below of his *perceptual* margin can be an incorrect approximation from below of his *perceptual and inferential* margin. A nice feature of our model is that these differences in margins are not stipulated.[16] They are accounted for by the epistemic updates themselves, since the margin after reflection is nothing but the margin in the updated model.[17]

4.2 Discounted margins

Epistemic updates help explain why and where things go wrong. But they can also be used to get positive results. Williamson assumes that Mr Magoo's estimate remains constant throughout the reflection process. This assumption is disputable, however. On a more realistic scenario, Mr Magoo's estimate of his current margin for error would become lower and lower as he goes through more and more rounds of reflections (intuitively, he is less and less sure about his margin, since many reflections have pushed it down). Clearly, such assumptions impact the soundness of the argument. To study exactly how, we need to take the values of Mr Magoo's successive reflections as parameters.

Let us consider *sequences of estimates* of the form $\overrightarrow{\eta} = \eta_1, \eta_2, ..., \eta_n, ...$ where η_n is Mr Magoo's estimate of his margin after the first $n-1$ rounds of reflection. An infinite sequence of iterations of the basic argument is thus determined. They are the formal rendering of the (potentially infinite) reflective process Mr Magoo engages in. Now the question is: under which conditions on $\overrightarrow{\eta}$ is it fine for us to freely iterate the argument? Or equivalently, under which conditions on $\overrightarrow{\eta}$ does Mr Magoo keep on learning things?

We are asking for a characterization of soundness conditions for an arbitrary number of iterations of the argument. A question about soundness only makes sense when a particular instance of the scenario has been chosen. So we fix d, Mr Magoo's perceptual margin for error. We shall say that a sequence of estimates $\overrightarrow{\eta}$ is *d-bounded* iff every partial sum is smaller than d, that is $\sum_{i=0}^{n} \eta_i \leq d$ for all n. We shall

[15] By *perceptual and inferential* knowledge we mean here the knowledge Mr Magoo gets from what he sees and from reflecting for the first time on the limitations of his perceptive abilities. Inferential knowledge itself comes in various degrees since Mr Magoo can then make inferences grounded in his first level perceptual and inferential knowledge.

[16] On top of the use of a unified knowledge operator, this is the second advantage of our approach over the one by Dokic and Égré in [5].

[17] Note that strictly speaking the updated model is not a margin model for the Euclidean topology on the reals. This is only because $\forall x\ ME(x,\eta)$ is asymmetric. To regain symmetry, and to get margin models as update models, we would need a stronger principle such as $\forall x\ (((s \geq (x-\eta)) \rightarrow \neg K(s < k)) \wedge ((s \leq (x+\eta)) \rightarrow \neg K(s > x)))$. Since nothing important hinges on this, we stick to the weaker principle. See [6] for a discussion of general margin for error principles in modal and epistemic logic.

also say that a sequence $\overrightarrow{\eta}$ makes Mr Magoo's argument *iteratively sound* iff every iteration of the argument starting from some premise $K(s < k)$ true in the actual world and with margin estimates chosen according to $\overrightarrow{\eta}$ is sound.[18]

Fact 7 *Let d be the margin of the margin model. A sequence of estimates* $\overrightarrow{\eta}$ *makes Mr Magoo's argument iteratively sound iff* $\overrightarrow{\eta}$ *is d-bounded.*

This fact should come as no surprise. Intuitively, the one shot version of the argument is sound if and only if Mr Magoo's estimate does not exceed his actual margin for error. Now, Mr Magoo might go for a sequence of cautious consecutive estimates instead of a one shot daring estimate. But in any case, he should not be allowed to outrange his initial perceptual margin for error. Which is just to say that consecutive estimates should not add up to more than d – the bounding condition on $\overrightarrow{\eta}$. What if Mr Magoo is overconfident? At some point in the reflective process, his estimates add up to more than his perceptual margin and he ends up with a false belief about the size of the tree. In terms of our iterated argument, the conclusion that he knows the size of the tree to be less than his initial estimate minus further improvements ends upon being false. This happens when the premise about the margin left at this stage is false.

Williamson considers that the margin estimate can be assumed to remain constant, if it is taken to be small enough. Fact 7 shows that this is too strong an assumption. If $\overrightarrow{\eta}$ is of the form $\eta_1, \eta_1, \eta_1 ...$, for some non-zero η_1, there is no d such that $\overrightarrow{\eta}$ is d-bounded. So no matter what the exact situation is, Mr Magoo is going to reach a false conclusion. By contrast, more realistic choices for $\overrightarrow{\eta}$ can yield iterative soundness. For example, if each estimate is no greater than it should be (considering the remaining margin at the current stage), that is if $\eta_n < d - \sum_{i=0}^{n-1} \eta_i$, then $\overrightarrow{\eta}$ is d-bounded (actually this is equivalent to $\overrightarrow{\eta}$ being d-bounded). This suggests an easy generalization of Fact 7. Let us say that a sequence of estimates $\overrightarrow{\eta}$ makes Mr Magoo's argument *iteratively coherent* if there is some situation (that is some margin model) such that the argument is iteratively sound in that situation. We get:

Fact 8 *A sequence of estimates* $\overrightarrow{\eta}$ *makes Mr Magoo's argument iteratively coherent iff* $\overrightarrow{\eta}$ *is d-bounded for some d.*

In this perspective, Williamson's account of Mr Magoo's scenario is merely wrong – there is no situation in which his assumptions about the ways of reflection yield an argument which remains sound when it is iterated. Using Quine's taxonomy (see [10]), the Margin for Error paradox may be classified as a *falsidical* paradox rather than as an antinomy. The false premise in this case is the assumption of constancy

[18] The precise definition of an iteration of the argument as parameterized by $\overrightarrow{\eta}$ and the proof of Fact 7 are given in the Appendix. Fact 7 might seem quite obvious, and indeed the proof is by no means difficult. Concluding the analysis of a paradox with a commonsensical conclusion may not be a bad thing, and one should keep in mind that the claims and proofs need to be done in CDEL instead of DEL.

of the margin estimates. Why does this falsidical paradox tend to look like an antinomy? We have the strong intuition that some sequences of estimates yield an iteratively coherent argument. This is indeed true. Because Williamson insists on the fixed estimate for the margin being arbitrarily small, we are misled into thinking that his choice is one of the choices yielding an iteratively coherent argument. In that case, we would have had an antinomy. But we do not.

4.3 The Surprise Examination

To conclude, we would like to briefly compare our solution with the thorough dynamic analysis of the Surprise Examination paradox due to Gerbrandy in [7]. The story is as follows. A teacher announces to her class on Monday that they will have a surprise examination this week. Clever Marilyn, a student in the class, starts thinking about it. First, the exam cannot be on Friday, because it would be known on Thursday evening that it will take place on Friday. Since it cannot be on Friday, it cannot be on Thursday either, because it would be known on Wednesday evening that it will take place on Thursday. By repeating the argument the student can conclude that there can be no surprise exam, which sounds like a plain contradiction. Let S be the statement that, for every day X in the week, if the exam is on day X, then the student does not know that it is on day X. Using epistemic updates, Gerbrandy shows that "If S correctly paraphrases the teacher's announcement, then Marylin's reasoning is cut short after having excluded the last day as the day of the exam" (p. 27). This is because there is no guarantee that S is true after the teacher has announced it (if the exam is on Thursday, S is true but $\langle S \rangle S$ is false). Gerbrandy suggests various ways to strengthen the teacher's statement. The teacher could explicitly say that the exam will still be a surprise after she has announced that it is a surprise. $S \wedge [S]S$ gives Marilyn enough information to exclude the last two days of the week, but it is still the case that $S \wedge [S]S$ need not be true after it has been announced, so that the exam could be scheduled on Wednesday. One might wish to go for an even stronger announcement δ that would state its own success, so that the equation $\delta \leftrightarrow S \wedge \langle \delta \rangle S$ holds. Gerbrandy shows that no formula of dynamic epistemic logic satisfies this equation, and that δ would be 'contingently paradoxical', in the sense that it would be both true and false in some situations.

Here is one way to look at the structural similarity of the Surprise Examination paradox and the Margin for Error paradox. The teacher's announcement that the exam will be a surprise allows the student to eliminate a day of the week as the day of the exam. After that, the announcement is not guaranteed to be true. If the teacher repeats her announcement, it is indeed bound to be false at some stage. Similarly, realizing that $ME(k, \eta)$ holds allows Mr Magoo to eliminate a certain range of heights as the height of the tree. After that, the principle is not guaranteed to be true. If Mr Magoo keeps on updating his knowledge according to the same estimate, the estimate will indeed be inaccurate at some stage. However, the Surprise Examination paradox involves a discrete scale, the consecutive days of the week, whereas the Margin for Error paradox involves a dense scale, the possible heights of the tree. This makes for a more subtle status of iterations. The fact that δ is (contingently) paradoxical essentially says that arbitrarily many iterations are not sound. Arbitrarily many iterations with a fixed estimate are not sound either, but Fact 7 states the conditions under which arbitrarily many iterations with a variable estimate are (iteratively) sound. Thus, in the somewhat richer setting offered by Williamson, epistemic updates can be applied to yield more fine-grained results concerning the demarcation line between paradoxical and non-paradoxical scenarios.[19]

5 Conclusion

We are finite creatures, endowed with limited perceptual abilities. Our knowledge obeys a margin for error. But we can explore and push our limits to some extent. What is the lesson to draw from Mr Magoo's story in this perspective? Well, there is good news and there is bad news. Here is the bad news first: properly speaking, we cannot *know* our limits. Knowing (an approximation of) our margin for error would make the notion either vacuous, or inconsistent. Just as in the case of Moorean sentences, our limitation is something we can *realize*, but this is not something stable for us to know. The good news is we can always improve on our limits. As long as our sequence of estimations is adequate, there is room for further improvement, and Fact 8 characterizes the conditions under which this may happen.

The attention given to the dynamics of knowledge is an essential component of our account. In this respect, we have followed the path opened by van Benthem [1] and taken up by Gerbrandy [7]. One novelty here is our use of Centered Dynamic Epistemic Logic instead of plain DEL, in a context in which DEL cannot satisfactorily handle the intended models for the paradoxical scenarios. We also put forward a notion of iterative soundness, in order to tease apart paradoxical and non-paradoxical versions of Williamson's scenario, in particular to distinguish between one application of Williamson's premises, and their iteration. In this respect, Williamson's paradox is a genuine sorites, since the core question is whether – or when – it is fine to repeat the argument. Because of that, we may wonder whether the dynamic approach may be extended to deal with other sorites more generally.

Acknowledgements We are grateful to O. Roy for detailed comments. The first author also wishes to thank the participants in the Logic and Language Seminar in Stockholm and in the Logic Semi-

[19] In his discussion of the Surprise Examination paradox, Williamson [12] considers a whole range of closely related paradoxes, so as to gradually turn the Margin for Error paradox into the Surprise examination paradox (to get started, think of a scenario in which Marilyn has a glimpse of a calendar on which the teacher has ringed the examination date, so that she knows it does not take place on Friday). According to Williamson, it is not exactly introspection which is to blame in the Surprise Examination paradox, but a similar assumption with respect to ascriptions of iterated knowledge. Williamson is certainly right that the similarities between the two paradoxes call for similar solutions. In this respect, the similarity between our solution to the Margin for Error paradox and Gerbrandy's solution to the Surprise Examination paradox is quite welcome.

nar in Gothenburg. We acknowledge the ANR project 'Cognitive Origins of Vagueness' (ANR-07-JCJC-0070) for financial support.

Appendix

A. Completeness proof for CDEL

We recall the main Theorem of Section 2 and give the proof for **K45** and the class of all frames. The proofs for **S5** and the class of reflexive frames would be similar. We extend the strategy used in [3] to capture epistemic updates as well as K operators.

Theorem 1 **K45** *(resp.* **S5***) plus the recursion axioms is sound and complete with respect to Centered Semantics with updates on the class of all frames (resp. of all reflexive frames).*

Proof. *Soundness*. We only prove soundness of the recursion axiom for knowledge. Soundness of **K45** can be proven along the lines of [3], and correctness for the first three recursion axioms is immediate. So for an arbitrary model $\mathscr{M}$ and a world w in $\mathscr{M}$, what we want is $\mathscr{M},w \vDash_{\mathrm{CS}} [\phi]K\psi$ iff $\mathscr{M},w \vDash_{\mathrm{CS}} \phi \to K(\phi \to [\phi]\psi)$. We get it easily by the following chain of equivalences:
$\mathscr{M},w \vDash_{\mathrm{CS}} [\phi]K\psi$
iff $\mathscr{M},(w,w) \vDash_{\mathrm{CS}} [\phi]K\psi$ (by definition of $\vDash_{\mathrm{CS}}$)
iff if $\mathscr{M},(w,w) \vDash_{\mathrm{CS}} \phi$, then $\mathscr{M}|\phi_w,(w,w) \vDash_{\mathrm{CS}} K\psi$ (by definition of $[-]$)
iff if $\mathscr{M},(w,w) \vDash_{\mathrm{CS}} \phi$, then for all $w' \in \mathscr{M}|\phi_w$ s.t. wRw',
$\mathscr{M}|\phi_w,(w,w') \vDash_{\mathrm{CS}} \psi$ (by definition of K)
iff if $\mathscr{M},(w,w) \vDash_{\mathrm{CS}} \phi$, then for all $w' \in \mathscr{M}$ s.t. wRw' and $\mathscr{M},(w,w') \vDash_{\mathrm{CS}} \phi$,
$\mathscr{M}|\phi_w,(w,w') \vDash_{\mathrm{CS}} \psi$ (by definition of $|\phi_w$)
iff if $\mathscr{M},(w,w) \vDash_{\mathrm{CS}} \phi$, then for all $w' \in \mathscr{M}$ s.t. wRw' and $\mathscr{M},(w,w') \vDash_{\mathrm{CS}} \phi$,
$\mathscr{M},(w,w') \vDash_{\mathrm{CS}} [\phi]\psi$ (by definition of $[-]$)
iff if $\mathscr{M},(w,w) \vDash_{\mathrm{CS}} \phi$, then for all $w' \in \mathscr{M}$ s.t. wRw',
$\mathscr{M},(w,w') \vDash_{\mathrm{CS}} \phi \to [\phi]\psi$
iff if $\mathscr{M},(w,w) \vDash_{\mathrm{CS}} \phi$, then $\mathscr{M},(w,w) \vDash_{\mathrm{CS}} K(\phi \to [\phi]\psi)$ (by definition of K)
iff $\mathscr{M},(w,w) \vDash_{\mathrm{CS}} \phi \to K(\phi \to [\phi]\psi)$

Completeness. The proof is by contraposition. Assume that a formula does not follow from **K45** plus the recursion axioms, we want to show that ϕ is not valid with respect to CS. We already know that **K45** plus the recursion axioms is complete with respect to standard semantics over the class of transitive and euclidian frames. So there is a model $\mathscr{M}$, based on such a frame, and a world w in it such that $\mathscr{M},w \nvDash \phi$. It is sufficient to show that $\mathscr{M},w \nvDash_{\mathrm{CS}} \phi$. This follows from a more general fact, namely that CS and Kripke semantics agree on transitive and euclidian models.

More precisely, for any transitive and euclidian model $\mathscr{M}$ and $w,w' \in \mathscr{M}$, for any formula ϕ in the language of dynamic epistemic logic, we have that $\mathscr{M},w' \vDash \phi$

iff $\mathscr{M},(w,w') \vDash_{CS} \phi$. The proof is by induction on the complexity of the formula. We only give the case for $\phi = [\psi]\chi$ (again, the other cases are taken care of in [3]).

By definition, we have that, on the side of Kripke semantics,
$\mathscr{M},w' \vDash [\psi]\chi$ iff if $\mathscr{M},w' \vDash \psi$, then $\mathscr{M}|\psi,w \vDash \chi$
and, on the side of centered semantics,
$\mathscr{M},(w,w') \vDash_{CS} [\psi]\chi$ iff if $\mathscr{M},(w,w') \vDash_{CS} \psi$, then $\mathscr{M}|\psi_w,(w,w') \vDash_{CS} \chi$

By induction hypothesis on ψ, we immediately have that $\mathscr{M},w' \vDash \psi$ iff $\mathscr{M},(w,w') \vDash_{CS} \psi$. So we just need to prove that $\mathscr{M}|\psi,w \vDash \chi$ iff $\mathscr{M}|\psi_w,(w,w') \vDash_{CS} \chi$. But the induction hypothesis on ψ also tells us that $\mathscr{M}|\psi$ are $\mathscr{M}|\psi_w$ are the same. Therefore, by induction hypothesis on χ, $\mathscr{M}|\psi,w \vDash \chi$ iff $\mathscr{M}|\psi_w,(w,w') \vDash_{CS} \chi$.

B. Positive Introspection as Dynamic Closure

In Section 3, we claimed that epistemic updates stand to our recasting of Williamson's argument as positive introspection stands to the original argument, modulo closure of knowledge under logical consequence and the recursion axioms. More precisely, the claim was:

Fact 2 $Kp \rightarrow KKp \vdash_{Cl,RA} Kp \rightarrow [Kp \rightarrow q]Kq$

Proof. First, we show:
$Kp \rightarrow [Kp \rightarrow q]Kq$
$\equiv_{RA} Kp \rightarrow ((Kp \rightarrow q) \rightarrow K((Kp \rightarrow q) \rightarrow q))$
By the recursion axiom for K,
$Kp \rightarrow [Kp \rightarrow q]Kq$
$\equiv_{RA} Kp \rightarrow ((Kp \rightarrow q) \rightarrow K((Kp \rightarrow q) \rightarrow [(Kp \rightarrow q)]q))$
Then by the recursion axiom for atoms:
$Kp \rightarrow ((Kp \rightarrow q) \rightarrow K((Kp \rightarrow q) \rightarrow [(Kp \rightarrow q)]q))$
$\equiv_{RA} Kp \rightarrow ((Kp \rightarrow q) \rightarrow K((Kp \rightarrow q) \rightarrow ((Kp \rightarrow q) \rightarrow q))$
Finally, $(Kp \rightarrow q) \rightarrow ((Kp \rightarrow q) \rightarrow q)$ is of the form $A \rightarrow (A \rightarrow B)$ which is equivalent to $A \rightarrow B$.

Now it is sufficient to show

$$Kp \rightarrow KKp \vdash_{Cl} Kp \rightarrow ((Kp \rightarrow q) \rightarrow K((Kp \rightarrow q) \rightarrow q))$$

Note that $Kp \rightarrow ((Kp \rightarrow q) \rightarrow q)$ is a tautology. But we have $Kp \rightarrow KKp$, so by closure under logical consequence $Kp \rightarrow K((Kp \rightarrow q) \rightarrow q)$. *A fortiori*, $Kp \rightarrow ((Kp \rightarrow q) \rightarrow K((Kp \rightarrow q) \rightarrow q))$.

Note that as may happen in DEL, the above entailment cannot be generalized to arbitrary formulae instead of the atoms.

C. Sequences of estimates

We recall the main Fact of Section 4, provide a precise characterization of what we mean by 'repeating the argument' and give a detailed proof. We work with margin models. In what follows d is Mr Magoo's perceptual margin for error and w_r is the actual world, so that epistemically possible worlds are worlds $w_{r'}$ with $|r - r'| \leq d$.

Fact 8 *A sequence of estimates $\overrightarrow{\eta}$ makes Mr Magoo's argument iteratively sound iff $\overrightarrow{\eta}$ is d-bounded.*

We start with some definitions. Let ζ_n be the sum of the first n margin estimates, that is $\zeta_n = \sum_{i=0}^{n} \eta_i$ and let $\langle ME(k, \overrightarrow{\eta})\rangle^n$ be short for the first n reflections, that is $\langle ME(k, \overrightarrow{\eta})\rangle^n = \langle ME(k, \eta_1)\rangle ... \langle ME(k - \zeta_{n-1}, \eta_n)\rangle$. In other words, $\langle ME(k, \overrightarrow{\eta})\rangle^n$ is the sequence of the first n updates according to Mr Magoo's reflective powers *as given by* $\overrightarrow{\eta}$. The premises at round $n+1$ are the same as at round 1, but for the fact that n consecutive reflections have taken place. So we get the premises for round $n+1$ essentially by prefixing $\langle ME(k, \overrightarrow{\eta})\rangle^n$ and adapting the parameters accordingly:

(3n+1') $\langle ME(k, \overrightarrow{\eta})\rangle^n K(s < k - \zeta_n)$
(3n+2') $\langle ME(k, \overrightarrow{\eta})\rangle^n ME(k - \zeta_n, \eta_{n+1})$
(3n+3') $\langle ME(k, \overrightarrow{\eta})\rangle^n (K(s < (k - \zeta_n)) \rightarrow [ME(k - \zeta_n, \eta_{n+1})] K(s < (k - \zeta_{n+1})))$

Proof. Assume that $K(s < k - \zeta_{n-1})$ is true at w_r, the net effect of $\langle ME(k - \zeta_{n-1}, \eta_n)\rangle$ at w_r is to delete all worlds $w_{r'}$ such that $r' \in [k - \zeta_n, k - \zeta_{n-1}]$, as can be proven by an easy induction on n. Let us now prove both directions.

($\Leftarrow$) If $\overrightarrow{\eta}$ is d-bounded, the argument is iteratively sound. We prove by induction on n that (3n+1'), (3n+2') and (3n+3') are true. First note that according to $\vDash_{CS}$, introspection always holds, so that (3n+3') is always true and we only need to check the first two premises. For $n = 1$, (1') is true by definition of super-soundness. Since $\overrightarrow{\eta}$ is d-bounded, we have in particular $\eta_1 \leq d$, so (2') is true[20]. For $n+1$, (3(n+1)+1') is true by induction hypothesis, since it follows from (3n+1'), (3n+2') and (3n+3') which are true. It only remains to be shown that $\langle ME(k, \overrightarrow{\eta})\rangle^n ME(k - \zeta_n, \eta_{n+1})$. It amounts to showing that when the space of world is restricted to $[0, k - \zeta_n[$, $ME(k - \zeta_n, \eta_{n+1})$ is still true in the actual world, that is $\mathscr{M}_d|[0, k - \zeta_n[, w_r \vDash_{CS} ME(k - \zeta_n, \eta_{n+1})$. We have just seen that $\mathscr{M}_d|[0, k - \zeta_n[, w_r \vDash_{CS} K(s < (k - \zeta_n))$, so we want $\mathscr{M}_d|[0, k - \zeta_n[, w_r \vDash_{CS} s < (k - \zeta_{n+1})$. Since η was d-bounded, we know that $\zeta_{n+1} \leq d$ and since we had $\mathscr{M}_d, w_r \vDash_{CS} K(s < k)$, $r < k - d$. Therefore $r < k - \zeta_{n+1}$ as needed.

($\Rightarrow$) If $\overrightarrow{\eta}$ is not d-bounded, the argument is not iteratively sound. By hypothesis, there is $n+1$ such that $\zeta_n \leq d$ and $\zeta_{n+1} > d$. Let k be such that $r + d < k < r + \zeta_{n+1}$. We have $\mathscr{M}_d, w_r \vDash_{CS} K(s < k)$. It is then sufficient to show that $\mathscr{M}_d|[k - \zeta_n, +\infty[, w_r \vDash_{CS} K(s < (k - \zeta_n)) \wedge (s \geq (k - \zeta_{n+1}))$. Since $\zeta_n \leq d$,

[20] See statement (i) on p. 13.

$\mathscr{M}_d|[k-\zeta_n,+\infty[,w_r \vDash_{\mathrm{CS}} K(s<(k-\zeta_n))$ follows from the proof for the other direction. We want $r \geq k-\zeta_{n+1}$. We know $k<r+\zeta_{n+1}$ hence $k-\zeta_{n+1}<r$.

References

1. van Benthem, J. (2004), "What One What One May Come to Know", *Analysis* 64 (282), 95-105.
2. van Benthem, J., Gerbrandy, J., Hoshi, T. and Pacuit, E. (2007) "Merging frameworks for interaction", *Journal of Philosophical Logic*, forthcoming.
3. Bonnay D. & Égré P. (2009), "Inexact Knowledge with Introspection", *Journal of Philosophical Logic*, 38 (2), pp. 179-228.
4. van Ditmarsch H., van der Hoek W. & Kooi B. (2007), *Dynamic Epistemic Logic*, Synthese Library, Springer.
5. Dokic J. & Égré P. (2009), "Margin for Error and the Transparency of Knowledge", *Synthese* 166, pp. 1-20
6. Égré P. (2008), "Reliability, Margin for Error and Self-Knowledge", in *New Waves in Epistemology*, D. Pritchard & V. Hendricks (eds.), Palgrave McMillan, pp. 215-250.
7. Gerbrandy, J. (2007) "The surprise examination paradox in dynamic epistemic logic", *Synthese*, 155, 21–33.
8. Gillies, A. (2001), "A new solution to Moore's paradox", *Philosophical Studies*, 105, 237-250.
9. Leitgeb, H. (2002), "Review of Timothy Williamson, *Knowledge and its Limits*", *Grazer Philosophische Studien*, 65: 195-205.
10. Quine W. V. O. (1961), "The Ways of Paradox", repr. in *The Ways of Paradox and Other Essays*, Harvard, third edition, pp. 1-18.
11. Williamson T. (1992), "Inexact Knowledge", *Mind*, 101, 217-42.
12. Williamson T. (2000), *Knowledge and its Limits*, Oxford University Press.