

Margins for Error in Context

Denis Bonnay
(DEC, ENS, Paris)

Paul Égré
(IJN, CNRS, Paris)*

According to the epistemic theory of vagueness defended in particular by Sorensen (2001) and Williamson (1994: 237), vagueness is due to our limited powers of discrimination: looking at a particular shade of red fabric, I may not be able to recognize that it is red, as a result of the specific granularity of my perceptual apparatus, which for instance makes the shade look somewhere between orange and red to me. Conversely, whenever I am confident that a particular shade of color is red, then this means that a slight variation in color should leave intact the fact that the shade is red. In Williamson's account of vagueness, this idea is expressed in terms of what Williamson calls *margin for error principles* for knowledge: whenever my knowledge is inexact in the sense of being approximative, it requires a sufficient margin for error in order to hold. More abstractly, the margin for error principle says that in order for me to know that some property P holds of an object d , then a slight modification of some relevant parameter in the object d should leave it in the extension of P . Expressed in terms of propositions and contexts, this means that in order to know that some proposition p holds in a context w , then p should still hold in a context w' that is only slightly different from w .

An important consequence of Williamson's margin for error semantics for inexact knowledge concerns the problem of iterations of knowledge. When applied to knowledge itself, the margin for error principle says that for me to know that I know p in a context w , I should know p in all contexts sufficiently similar to w . If knowledge is taken to be positively introspective, namely to be such that I know that I know p whenever I know p , then one can build a soritical argument to the effect that if I know p in a context w , then step by step, I should continue to know p even in contexts that are informationally very remote from w , including contexts in which p is false. For Williamson, the contradiction shows that knowledge, just like other mental states, is not luminous, namely that one can know p without knowing that one knows p .

In Dokic & Égré (2004), an axiomatic version of Williamson's argument was challenged, resting on the idea that higher-order knowledge and first-order knowledge need not obey the same margins of error. More recently, in Bonnay & Égré (2006), we showed how Williamson's margin for error semantics for knowledge can be modified in order to let the margin for error principle and the principle of positive introspection coexist. Despite their common inspiration, there remains a gap between the two approaches. On the one hand, as we shall see below, it can be shown that our semantic approach, like the one of Dokic & Égré, turns out to deny one of the premises of the syntactic version of Williamson's argument, namely the idea that the agent

*Version of 01/02/2007 - To appear in M. Garcia-Carpintero & M. Kölbel (eds), *Relative Truth*.

can have systematic knowledge of her margin of error. On the other hand, although both the syntactic approach followed by Dokic & Égré and the semantic approach put forward by Bonnay & Égré consist in adding a contextualist component to the evaluation of epistemic sentences, this common ingredient needs further articulation. In Bonnay and Égré (2006), in particular, we formulate a two-dimensional semantics for knowledge whose contextualist inspiration can be made more explicit if one introduces actuality operators. Moreover, the introduction of actuality operators will also allow us to handle the axiomatic version of Williamson's argument, while remaining closer to its initial premises. In what follows, our aim will therefore be to spell out the details of our contextualist approach to Williamson's paradox. The main point which we will develop concerns the idea that iterations of knowledge should remain anchored to the initial context of epistemic evaluation, in a way that blocks the soritical progression. In this respect, the paper should be seen as an effort to bring the two paradigms of epistemicism and contextualism about vagueness closer together.

The paper is structured as follows: in the first section we review the syntactic and semantic versions of Williamson's paradox; in section 2 we present a solution to the semantic version in the framework of Centered Semantics, a two-dimensional semantics for epistemic logic. We discuss in the following section the connection between Centered Semantics and actuality operators, and use this connection for the analysis of the syntactic version of the paradox. We conclude with some comparisons in section 4, in particular with Kamp's contextualist treatment of the sorites.

1 Margins and iterations

Margin for error principles and knowledge iterations do not live well together, as they give rise to paradoxical conclusions. Williamson was the first to observe this phenomenon, and turned it into an argument against the view that knowledge obeys the (KK) principle of positive introspection, namely that whenever I know p , I am in a position to know that I know p . Others, however, in particular Gomez-Torrente (1997), and then Fara (2002), have expressed worries about this, by pointing out that for some propositions at least, the (KK) principle seems clearly to be valid, thereby putting into question the validity of margin for error principles. In this opening section, we want to review two versions of Williamson's paradox, first a semantic version, linked to Williamson's margin for error semantics for knowledge (Williamson 1994), and then a syntactic version, as exposed by Williamson (2000). The two versions will be needed in order to evaluate the limitations of our own semantic approach to the paradox.

Williamson's paradox can be presented in more than one guise. Williamson himself has given several versions of it (1992, 1994, 2000): in all cases, the scenario is that of an agent who is to give an estimation about some quantity, or to express a qualitative judgment depending on the variation of some parameter. For instance, it can be a situation in which I observe a group of people, and wonder whether it makes a crowd, depending on how many people I see (Williamson 1994); or it could be a situation in which I have to make an explicit estimation about the height of a tree that I observe at some distance (Williamson 2000); or it can be a situation in which I make judgments about whether I feel cold or not as the temperature varies (Williamson 2000). The examples can easily be multiplied. As yet another example in what follows, we shall consider a

situation in which I observe a series of sticks at a certain distance, linearly ordered by size, and in which I am asked whether they fit in a certain box: in that case, I have to make a decision about the precise predicate “fits in the box”. The predicate “fits in the box” is precise in the sense that whether an object fits in the box or not really depends on objective features of the world, and not on specific properties of the language. Indeed, by contrast to a predicate like “bald”, “far”, or even “crowd”, the predicate does not seem semantically vague (namely such that it may fail to have a determinate extension)¹. We are more prone to admitting the existence of an objective cut-off point between the objects that fit in the box and those that do not, as assumed by the epistemic theory of vagueness more generally, than in the case of predicates like “bald” or “far”, for which at least some ingredient of semantic vagueness seems to play an important role. For a scenario of this kind, the intuition is therefore clear that the predicate “fits in the box” has a determinate extension, and that our inability to say whether it applies to this or that object is imputable to the limitations of our discriminative capacities (as an effect of distance and other visual parameters).

1.1 The paradox, semantically

The easiest way to present Williamson’s paradox is to start with the semantic version. Intuitively, to say that knowledge obeys a margin for error principle means that in order to know that some proposition ϕ holds in a context w , ϕ must hold in all contexts that are sufficiently similar to w . For instance, consider a situation in which I reflect upon whether I feel cold over time (Williamson 2000). Suppose that in order for me to know that I feel cold at time t , not only must I feel cold at t , but I must feel cold already at time $t - 1$, and also at time $t + 1$. This gives us a margin for error principle expressing the idea that in order for me to become aware that I feel cold, my feeling of cold must last sufficiently long, or be sufficiently intense. Likewise, consider the situation in which I observe a series of distant sticks, such that all and only sticks of size less than 4 inches will fit in the box. Seeing the sticks at a distance, my judgments about whether they fit in the box need a sufficient margin for error in order to be reliable. By analogy to the temporal case, we can suppose that as a fact of my perception, in order for me to know that a stick of size n fits in the box, a stick of size $n + 1$ or $n - 1$ inches also has to fit in the box.

Both scenarios can be represented by means of the same linear Kripke model, a particular case of what Williamson 1994 defines as a *fixed margin model*, in which we let the worlds be named by the natural numbers: in the temporal case, the points represent successive moments of time; in the box scenario, they represent the different sticks ordered by sizes; more generally, the model can be used to represent the possible values of some scalar quantity.

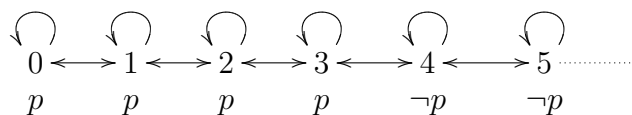


Figure 1: A margin for error model

¹See Kölbel (forthcoming) for a detailed characterization of semantic vagueness.

Two points i and j of the model are connected by an arrow if and only if they belong to the same interval of error, namely if $|i - j| \leq 1$. Letting p stand for “I feel cold”, then the model depicts a situation in which I feel cold from time 0 up to time 3, and start not to feel cold from time 4 onward (or indeed somewhere between 3 and 4). If instead p denotes the property of fitting in the box, then the model represents the fact that all and only sticks of size less or equal than 3 inches fit in the box. Like Williamson, we shall assume that the margin for error principle captures exactly the truth conditions of knowledge sentences. With respect to the model, this means that for every i :

(MS) $i \models K\phi$ if and only if for every j such that $|i - j| \leq 1$, $j \models \phi$

Suppose now that knowledge satisfies the (KK) principle of positive introspection. Then this means that for every i in the model:

(KK) If $i \models K\phi$ then $i \models KK\phi$.

It is now easy to see how a contradiction can be obtained from (KK) and (MS): suppose that $n \models Kp$: for instance, at time n , I know that I am cold; or of the n -th stick in the series, I know that it fits in the box. By (KK), this entails that I know that I know it, namely: $n \models KKp$. From this and (MS), it follows in particular that $n + 1 \models Kp$. Thus, if I know that I am cold at time n , then I also know that I am cold at time $n + 1$. By induction, it means that if I know p at n , then I must know p at all subsequent points. Thus, supposing that I know I am cold at time 0, then I know that I am cold at all later times, a contradiction if the temperature can rise. Likewise, from my knowledge that a stick of size 1 fits in the box, it should follow that I know that sticks of any size however large fit in the box as well, which is a contradiction.

Several remarks can be made on the assumptions we used. Firstly, as Gomez-Torrente (1997) originally pointed out, the full generality of the (KK) principle is not needed to reach the soritical conclusion. Indeed, it suffices to assume that there is some i such that $i \models Kp$, and for which all iterations of knowledge hold. Propositions p satisfying this property are called *transparent* by Fara (2002). Likewise, the assumption that (MS) captures *exactly* the truth-conditions of knowledge is not needed either to derive the paradox. The “if and only if” in (MS) can be weakened into an “only if”, and the K operator may be given separate truth conditions.² Also, it can be noted that we only made use of the “upward” version of the margin for error principle in order to get a contradiction. In the temporal case, in particular, it would have been enough to work with non-symmetric margins for error, that is to suppose that in order to know some proposition at time t , then p must hold at t and at the next time, but not necessarily at the previous time. Similarly, if we think of the sticks scenario, if I know that a stick of size n fits in the box, then presumably I should know of smaller sticks that they fit too, but not so easily for larger sticks. The reason to assume symmetric margins of error, however, is that the accessibility relation in the above margin model can be taken to represent the relation of perceptual indiscriminability, which is standardly assumed to be symmetric, the crucial feature being the fact that this relation is reflexive and non-transitive. Here we assume symmetric margins of error, as they also allow

²In the temporal case, for example, the upward version of the margin for error principle is expressible as $K\phi \rightarrow X\phi$, where X is the “next” operator. See Égré 2006 for a discussion of bimodal formulations of margin principles in terms of neighborhood modalities.

us to get “downward” versions of the epistemic sorites we presented, but the reader should bear in mind that nothing substantial rests on this assumption, and may just as well replace the model of Figure 1 with a reflexive, right-directed model (in which iRj if and only if $j = i$ or $j = i + 1$). Finally, it is easy to see the same argument we presented can be used to show that (MS) conflicts with the principle of negative introspection, which says that if I don’t know ϕ at t , then I know that I don’t know that ϕ at t . In the model of Figure 1, $3 \models \neg K\neg p$. Assuming negative introspection from this point onward, it is easy to see that $\neg K\neg p$ should hold at all subsequent points, thereby contradicting the fact that $5 \models K\neg p$.

1.2 The paradox, syntactically

Let us now turn to the syntactic version of the puzzle, as appears for instance in Williamson (1992, §1) and Williamson (2000, c. 5) : by syntactic, we have in mind the reconstruction of the paradox by means of explicit axioms and rules of inference. This approach contrasts with the semantic one in so far as we do not need to start from explicit truth-conditions for knowledge. We now consider a language consisting of atomic propositions p_i , with $i \in \mathbb{N}$, with an epistemic operator K and the usual rules of constructions for formulas. From an intuitive semantic point of view (which the reader should keep in mind), each proposition p_i should now be taken to express an indexical proposition, for instance: “the temperature is i degrees” (over some appropriate scale), or: “the stick is i inches long”. This makes another important difference with the approach of the previous section: instead of using the indices i in the metalanguage, we now introduce indices explicitly in the object language. Furthermore, where $i \models Kp$ meant “at time i /when the temperature is i degrees, I know that it is cold”, Kp_i now means: “I know that the temperature is i degrees”; likewise, where $i \models Kp$ meant “looking at a stick of size i , I know that it fits in the box”, Kp_i means: “the stick I am looking at is i inches long”. Consequently, the form of the puzzle we discuss now involves quantitative instead of qualitative propositions.

In the presentation of the syntactic version of the paradox, Williamson considers a scenario in which a subject reflects upon his perceptual limitations. In particular, the subject is now supposed to know that his knowledge involves a specific margin for error. In the case where the margin is of 1 unit, we get a reflective version of the margin principle (ME), which we call (KME) (for Knowledge of the Margin of Error); that is, we have for every $i \geq 0$ (letting $i - 1 = 0$ for $i = 0$):

$$(ME) \quad K\neg p_i \rightarrow (\neg p_{i-1} \wedge \neg p_i \wedge \neg p_{i+1})$$

$$(KME) \quad K[K\neg p_i \rightarrow (\neg p_{i-1} \wedge \neg p_i \wedge \neg p_{i+1})]$$

The principle (ME) says that in order for me to know that the stick I am looking at is not i inches long, then the stick can’t be $i + 1$ or $i - 1$ inches either; its iterative version, (KME), states a constraint of rationality on the part of the agent. Not only is the agent’s knowledge subject to a margin for error, but the subject is aware that his knowledge obeys a margin for error principle. As a second rationality constraint, implicit in the semantic version of the puzzle, the knowledge of the subject is supposed to be closed under logical consequence, as expressed by the following axiom (K) (for Kripke’s axiom):

$$(K) \quad K(\phi \rightarrow \psi) \rightarrow (K\phi \rightarrow K\psi)$$

We suppose moreover that the subject’s knowledge is positively introspective:

$$(KK) \quad K\phi \rightarrow KK\phi$$

With the rule of modus ponens and the rule of substitution, the three principles (KK), (KME) and (K) are sufficient to replicate the foregoing soritic argument. Indeed, suppose now that $K\neg p_n$, that is the subject knows that the stick he is looking at is *not* of size n . By (KK), he knows that he knows it, $KK\neg p_n$. From (KME) and (KK), it holds in particular that $K(K\neg p_n \rightarrow \neg p_{n+1})$. From (K) and the last two formulas, it follows that $K\neg p_{n+1}$. Thus, from my knowledge that the stick is not of size n , I know that it is not of size $n + 1$. By induction, this means that for every n , I know that the stick is not of that size, a plain contradiction if the stick is supposed to be of a determinate finite size.

The syntactic version of Williamson’s puzzle is slightly more involved than the semantic one, which only involved the two principles (MS) and (KK). The two versions of the puzzles are easily related, however: first, it is easy to see that the (K) principle is true at every point of the model of Figure 1, as a consequence of the semantic clause expressed by (MS).³ Likewise, if one assigns to every atom p_i the value true at i , and the value false everywhere else, as expected from the intended meaning of p_i , one can check in the same way that (KME) – like (ME) – is also universally true in the model.⁴ By contrast, (KK) is not valid in the model, since for instance $2 \models Kp$, but $2 \not\models KKp$. Thus, the two principles (K) and (KME), which may seem to express *prima facie* stronger principles of rationality than (ME) and (KK) themselves, are already validated by Williamson’s margin for error semantics. Intuitively, however, (KME) does correspond to a more demanding condition than its non-reflective version (ME). The semantics we present in the next section will make that intuition precise.

2 Centered semantics

In the previous section we presented two versions of Williamson’s paradox: first a semantic version involving qualitative predicates, then an axiomatic version involving quantitative predicates. In each case we get a paradox in the sense that, starting from plausible premises and plausible rules of inference, we end up with an obviously false conclusion. For Williamson, the paradox suggests that one of the premises must be rejected, namely the (KK) principle. In this section we present a different solution: in our view, neither (KK) nor (MS) is to be rejected strictly speaking. Rather, we make a more holistic move and propose a modification of the underlying semantic framework: roughly, the idea is that epistemic sentences ought to be evaluated with respect to more than one index in order to avoid spurious dependencies and to get a more realistic model of the agent’s epistemic situation. Before introducing our semantics, it will be useful to review the general features of the standard epistemic semantics.

³Clearly, if $i \models K(\phi \rightarrow \psi)$ and $i \models K\phi$, then this means that at every j such that $|i - j| \leq 1$, $j \models \phi \rightarrow \psi$, and likewise $j \models \phi$, so $j \models \psi$, and thus $i \models K\psi$.

⁴Suppose that $j \models K\neg p_i$; by (MS), this means that for every n such that $|j - n| \leq 1$, $n \models \neg p_i$, that is $j \neq i + 1$, $j \neq i$ and $j \neq i - 1$, and so $j \models \neg p_i \wedge \neg p_i \wedge \neg p_{i+1}$. So every j satisfies $K\neg p_i \rightarrow (\neg p_{i-1} \wedge \neg p_i \wedge \neg p_{i+1})$: this holds relative to every point k such that $|k - j| \leq 1$, so $K[K\neg p_i \rightarrow (\neg p_{i-1} \wedge \neg p_i \wedge \neg p_{i+1})]$ also holds everywhere in the model.

2.1 Iterations of knowledge and context-shifting

According to the standard Kripke-Hintikka semantics for knowledge, I know p in a context w if and only if p holds in all circumstances that are compatible with the information available to me at w (including the actual world, since knowledge is true). Standardly, a situation of knowledge is represented by means of a Kripke model $M = \langle W, R, V \rangle$, where W is a set of epistemic states or contexts, R is the epistemic accessibility relation, and V is a function specifying which atomic sentences are true at which worlds. Williamson’s margin for error semantics for knowledge, as expressed for instance by clause (MS), is only a particular case of the standard semantics, according to which, given such a model M , $M, i \models K\phi$ iff for every $j \in W$ such that iRj , $M, j \models \phi$.

An important consequence of the standard semantics concerns sentences with iterated modalities: to check whether a sentence of the form $KK\phi$ holds at i , one needs to check whether ϕ holds at worlds that are two steps away from i along R in the model. More generally, for a sentence with n nested modalities, one has to visit worlds that are n -steps away from i along the accessibility relation. In the model of Figure 1, for instance, $1 \models KKp$, but $1 \not\models KKKp$, since 4 is a world accessible in three steps in which p is not true. To put it otherwise, whether $1 \models KKKp$ depends on whether $3 \models Kp$: however, it is arguable that whether $1 \models KKKp$ should depend only on whether $1 \models Kp$, and thus only on perceptual alternatives to context 1, not on perceptual alternatives to context 3. With regard to the standard semantics, we may therefore say that iterations of knowledge shift the context, a phenomenon that is crucial to the generation of the soritical conclusion when positive introspection is supposed to be universally true in the model.⁵

This feature of the standard semantics is precisely what we think should be challenged to handle Williamson’s paradox. In a margin model like that of Figure 1, the relation of epistemic accessibility can be taken to represent the notion of perceptual indiscriminability. As such the relation represents the fact that I cannot perceptually discriminate between pairwise adjacent sticks, but that I can discriminate between pairwise non-adjacent sticks. Thus, in order to know that a given stick fits in the box, all sticks that are adjacent to it must fit too. The question then is to what extent higher-order levels of knowledge have to depend on perceptual alternatives that are more remote than those needed to secure knowledge at the first level. For Williamson, such long-distance dependencies give an accurate picture of the attainability of higher-order knowledge. Indeed, Williamson sees iterations of knowledge as a process of “gradual erosion”, precisely

⁵Note that our use of the notion of context-shifting is consistent with the Kaplanian distinction between circumstance-shifting operators and context-shifting operators (See Kaplan 1989, and Recanati 2000 on the distinction). Standardly, “know” is described as circumstance-shifting because the truth of “ X knows that p ” in a context w depends on the truth of p in all circumstances or epistemic alternatives compatible with w for X . Precisely, we agree that K is circumstance-shifting, but it should not be context-shifting (when iterated). Our claim is that the truth of iterated knowledge modalities in a context w should depend only on the circumstances that are relative to w , not on circumstances that are relative to contexts other than w , as they seem to do over non-transitive models. In S5 models of knowledge in which the accessibility relation is an equivalence relation, the accessibility relation is transitive, and this phenomenon of context-shifting does not arise. More abstractly, therefore, what we want and achieve in the next section can be described as a way of validating positive introspection over non-transitive models (Bonnay & Égré 2006).

because, according to him, iterations of knowledge “widen margins of error”.

We disagree with this conception. The context-shifting phenomenon is related to what we consider to be three main problems for this view. The first is a supervenience problem for higher-order knowledge: Williamson considers that higher-order knowledge supervenes on information in the world that lies beyond the information on which first-order knowledge supervenes. In our view, however, higher-order knowledge depends only on whether first-order knowledge occurs, without requiring the subject to gather more information in the world. The second problem concerns the computation of higher-order knowledge: Williamson is committed to the view that, depending on the context, I may know that I know...that I know p , and still fail to know that I know that I know...that I know p . From a cognitive point of view, this does not seem very plausible. Indeed, we become very quickly unable to keep track of levels of knowledge and it is even hard to say what a sentence like “Peter knows that he knows that it’s raining, but does not know that he knows that he knows” really means. The third problem is a problem of modularity: Williamson’s account presupposes that higher-order levels of knowledge are subject to the same margins of error that apply at the first-order level. This supposition does not speak for itself, however, especially if we take perceptual knowledge and the reflective knowledge on one’s perceptual knowledge to be two different modes of knowledge. In the next section we propose an alternative picture of higher-order knowledge, in which higher-order knowledge and first-order knowledge supervene on the same informational facts, in such a way that if I know p , I thereby know that I know p , without this second-order knowledge introducing an additional margin of error. This picture of higher-order knowledge is essentially the Cartesian picture, in which if I know something in a context w , then my knowing that I know remains anchored to the same context. Building on the previous intuitions, our aim is to show that the Cartesian picture is at least coherent and compatible with the idea that first-order knowledge comes about with a margin of error.

Before examining the details, however, it is fair to say that Williamson’s own criticism of the Cartesian view gains much plausibility when the operator under discussion is the operator “it is clear that”, which is the operator Williamson originally considers in his account of vagueness (Williamson 1994, Appendix). A central aspect of the vagueness phenomenon, in Williamson’s account, concerns higher-order vagueness. Higher-order vagueness corresponds to the idea that some property can be clearly exemplified, without it being clear that the property is clearly exemplified: for instance, some object may clearly be red, without it being clear that the object is clearly red. Likewise, an object may not clearly be red, without it being clear either that the object is not clearly red.⁶ In the picture of higher-order knowledge that we defend, by contrast, if I know that the pen fits in the box, then I know that I know that the pen fits in the box. Our account also validates that if I don’t know whether the pen fits in the box, then I also know that I don’t know. The problem then is to determine to what extent the operators “it is clear (to me) that” and “I know that” can be considered equivalent in a situation of inexact knowledge like the one we are discussing.

⁶We actually perceive an asymmetry between the two judgments: “if it is clear that p , then it is clear that it is clear that p ” (schema 4), and “if it is not clear that p , then it is clear that it is not clear that p ” (schema 5). In our view, the failure of schema 5 is much more intuitive than that of schema 4, and gives a better illustration of higher-order vagueness.

Our response to this problem is that the two operators are most likely substitutable at the first-level, but not at higher-order levels. Certainly, if it is not clear to me that the pen fits in the box, then I don't know that it fits. Arguably, in the opposite direction, if it is clear that the pen fits in the box, then I know that it fits in the box. Is it contradictory, however, to suppose that it is not quite clear that it is not clear whether the pen fits in the box, and yet to say that I know that it is not clear whether the pen fits in the box? We don't see a contradiction here: if it is not clear whether the pen fits, and still not quite clear that it is not clear, I can perfectly reflect on my situation and be aware that it is not clear whether the pen fits or not, and yet, precisely because it is not clear that it is not clear, think that probably little would be needed for me to know whether or not the pen fits. In this regard, we agree with Halpern (2004) that a distinction should be made between the operators "it is clear that" and "I know that". More precisely, what we say here is that even if they are taken to be synonymous in front of atomic sentences, they should no longer be taken to be such when iterated.

Consequently, we think it makes sense to preserve introspection principles such as "If I know that p , I know that I know p ", and even "If I don't know that p , I know that I don't know", without thereby committing ourselves to the denial of higher-order vagueness. By saying this, we consider that second-order knowledge describes a form of categorical awareness of one's first-order epistemic state, more than an iterated judgment of clarity on content of that state proper.

2.2 Fixing the context

To avoid the context-shifting phenomenon induced by iterations of knowledge, we present a modification of the standard Hintikka-Kripke semantics, which we call Centered Semantics (Bonnay & Égré 2006), in which epistemic sentences are evaluated with respect to two indices. The semantics, as we shall see below, bears some interesting connections to two-dimensional formalisms developed earlier for temporal and for epistemic logic (Kamp 1971, Rabinowicz & Segerberg 1999, Halpern 2004). As just explained, the original motivation for the present approach was primarily to capture a notion of modularity for knowledge, however, namely to allow first-order knowledge and higher-order knowledge to obey distinct constraints (Égré 2006), and moreover to have a plausible model of knowledge iterations in situations of bounded rationality (Égré & Bonnay 2006).⁷

At the level of indices, this idea of modularity is cashed out in the following way: given a couple (w, w') , the second index w' determines the atomic information available at w' ; the first index w , by contrast, determines what the subject knows, including about himself, when he is located at w . Epistemic sentences are evaluated with respect to couples (w, w') such that w' is

⁷CS is a particular case of a family of resource sensitive semantics, which we call Token Semantics, of which the standard Kripke semantics is also a particular case (Bonnay & Egré 2006). We only focus on CS in the present paper, first because CS is the first semantics in the hierarchy of Token Semantics (with Kripke semantics at the other extreme), but also because it is the most appropriate framework for the discussion of the issues of epistemic indexicality connected to Williamson's paradox. One feature of TS that is relevant for the problem of introspection concerns the idea that trivialization of knowledge iterations need not occur at level 2, as happens with CS, but only at higher levels.

a world perceptually accessible from w : in this way, the index w serves to fix the perspective of the subject. Borrowing from the terminology of Rabinowicz and Segerberg 1994, we may call the first index the “perspective-world” and the second index the “reference-world”. Typically, however, we are interested in what the subject knows at w , when the primary information he receives comes from w itself. Hence, we define the semantics in two stages: first relative to arbitrary couples (w, w') , then relative to single indices, corresponding to the diagonal case in which $w = w'$. Given a Kripke model $\mathcal{M} = \langle W, R, V \rangle$, the satisfaction is defined inductively as follows:

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

According to clauses (i)-(iii), atomic sentences and boolean compounds thereof depend on the second index only, namely the reference-world; by contrast, the evaluation of epistemic operators depends essentially on the first index or perspective-world, as expressed by clause (iv). This feature becomes crucial for the evaluation of sentences with iterated modalities: in that case, clause (iv) ensures that we keep track of the initial point of evaluation, seen as the “center” of the worlds that count as relevant alternatives in the evaluation of epistemic modalities.

If we consider the model of Figure 1, we can check indeed that the semantics is equivalent to the standard semantics for propositional sentences and for sentences without nested modalities. For instance, $2 \models_{CS} Kp$, just as in the standard case, since $(2, 1), (2, 2),$ and $(2, 3) \models_{CS} p$, and 1, 2, 3 are the only worlds accessible from 2. The situation changes for the case of iterated modalities: with the standard semantics, $2 \not\models KKp$; however, it is easy to check that $2 \models_{CS} KKp$: by definition, $(2, 2) \models_{CS} KKp$ if and only if $(2, 1), (2, 2)$ and $(2, 3) \models_{CS} Kp$, if and only if all of these in turn satisfy p , which we have seen to be the case.

More generally, relative to a world w , the semantics bounds the number of possible moves in the model to those worlds that are accessible from w in one step. In this manner, whether KKp holds or not at a world w depends only on whether Kp already holds *there*, and not on whether Kp holds at more remote indices. In centered semantics, iterations of knowledge no longer give rise to context-shifting, but remain anchored to the context relative to which knowledge of the first-order is determined. Thus, in the same way in which it articulates a notion of modularity, centered semantics also captures an element of indexicality of higher-order knowledge.

Before saying more about this notion, it is now easy to see how the semantics blocks Williamson’s paradox. First, the reader can check that the semantics validates principle (KK) of positive introspection, that is $w \models_{CS} K\phi$ implies $w \models_{CS} KK\phi$. In Bonnay & Egré (2006), it is shown that the semantics is sound and complete for the normal logic K45, namely for the axiom system which validates, besides axiom K and (KK), the principle of negative introspection (namely $\neg K\phi \rightarrow K\neg K\phi$). If we graft the semantics directly onto Williamson’s fixed margin semantics, which validates the axiom T of veridicity ($K\phi \rightarrow \phi$), the resulting semantics becomes sound and complete for the system S5 of fully introspective knowledge. Thus, if we consider

(MS) above in section 1.1, its formulation in the framework of centered semantics becomes:

(MS') $i \models_{\text{CS}} K\phi$ iff for every j such that $|i - j| \leq 1$, $(i, j) \models_{\text{CS}} \phi$

namely: I know ϕ in context i if and only if ϕ holds at every world j that is perceptually indistinguishable from i . With this enrichment of the semantics, the principles of positive introspection and of margin for error are compatible again. In particular, from $n \models_{\text{CS}} Kp$, it follows that $n \models_{\text{CS}} KKp$, since positive introspection is CS-valid. By (MS'), this entails in particular that $(n, n + 1) \models_{\text{CS}} Kp$. But from that, it does not follow that $n + 1 \models_{\text{CS}} Kp$, since this would mean that $(n + 1, n + 1) \models_{\text{CS}} Kp$. Another way to see the point is by considering the above model: in the model, $Kp \rightarrow KKp$ is CS-true at every point of the model, likewise Kp holds at point 0, but now this is compatible with the fact that p holds only in a limited segment of the model. As a consequence, the move from the standard semantics to centered semantics blocks the soritic progression. Despite this, (MS') remains as close as possible to (MS): indeed, (MS') still means that in order to know some proposition in a given context, that proposition must be true in all contexts that are sufficiently similar to it; but this dependence to the initial context is now more tightly articulated.

3 Knowledge and actuality

Centered semantics makes margin for error models and introspection principles compatible, and blocks the path to Williamson's paradoxical conclusion. What about the syntactic version? Centered semantics, as we shall see, turns out to invalidate one of the premises of Williamson's argument, namely principle (KME), by which the agent knows her margin of error. The rejection of that premise, however, is arguably too strong: the aim of this section is to examine whether this is so, and to give a more precise discussion of the status of this principle. To do so, we present a reformulation of Centered semantics in terms of actuality operators. The idea that perceptual possibilities are indexical – what I can or cannot perceive only depends on what is indistinguishable from the actual situation I am in – is indeed hardwired in our semantics. The introduction of actuality operators allows us to make explicit this implicit indexical element, and to articulate the semantic and syntactic versions of the paradox more precisely.

3.1 Knowing the margin

Williamson's syntactic paradox shows the inconsistency of a set of four principles: in order to escape the contradiction, and short of changing the logic, it is necessary to reject at least one of the principles used in the argument. Given the correspondence between the syntactic and the semantic versions of the paradox, we must expect centered semantics to invalidate at least one of those principles. This can't be (KK), since centered semantics was designed to make it valid. This should not be (ME), since the whole point of (ME) is to account for the inexactness of perceptual knowledge. Hence there are only two principles left for rejection, namely (K) and (KME).

Centered semantics does validate (K), and we think this is at it should be. (K) expresses closure of knowledge under logical inferences: if I know that p implies q and if I also know

that p , then I know that q . As a rationality requirement, (K) is an idealization. Real life agents are not logically omniscient, they do not consciously know all the consequences of what they consciously know, and in this respect the principle is obviously very strong. For our purposes, however, rejecting (K) would not count as a real solution to the syntactic paradox: for the paradox could still be derived for a logically omniscient agent with inexact knowledge (even brilliant mathematicians get cold sometimes).

So what about (ME) and (KME)? Let us state these principles in their syntactic form as in paragraph 1.2:

$$(ME) \quad K\neg p_i \rightarrow (\neg p_{i-1} \wedge \neg p_i \wedge \neg p_{i+1})$$

$$(KME) \quad K[K\neg p_i \rightarrow (\neg p_{i-1} \wedge \neg p_i \wedge \neg p_{i+1})]$$

Until now we have used a qualitative margin for error model, in which a given proposition like “I feel cold” or “the pen fits in the box” held up to some threshold. In order to match the syntactic formulation of (ME) and (KME), we can easily turn this model into a quantitative model, in which atomic propositions like p_3 : “the pen is 3 inches tall” express the underlying quantities:

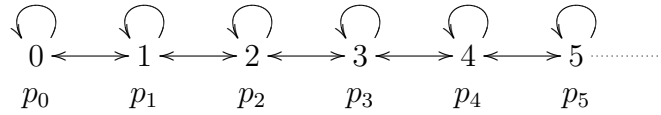


Figure 2: Margin model with quantitative propositions

(ME) is valid on this model. For example, $4 \models_{CS} K\neg p_2$. For (ME) to hold, it must hold that $4 \models_{CS} \neg p_1 \wedge \neg p_2 \wedge \neg p_3$, which is the case. By forcing here that $4 \models_{CS} \neg p_3$, (ME) implements my margin for error: if I was looking at a pen of size 3, I would not be able to know that it is not of size 2, because my eyes cannot distinguish two pens whose sizes differ by no more than one inch.

However, surprisingly enough, the validity of (ME) on a model does not imply the validity of (KME) on that same model. To see this, look at what happens at point 4 in the model of Figure 2. For $K(K\neg p_2 \rightarrow \neg p_3)$ to hold at 4, we must have $(4, 3) \models_{CS} K\neg p_2 \rightarrow \neg p_3$. Clearly, $\neg p_3$ is false at $(4, 3)$, because $(4, 3)$ is the perceptual alternative to 4 in which the pen is of size 3. But $K\neg p_2$ is true at $(4, 3)$, just because $\neg p_2$ is true at all perceptual alternatives relevant to the situation 4 in which I am in, namely $(4, 3)$, $(4, 4)$ and $(4, 5)$.

To make things clear, it will be useful to define the following notions in full generality. We refer to Kripke semantics as KS in what follows.

Definition 1. Let $\mathcal{M} = \langle W, R, V \rangle$ be a Kripke model, S a semantics, a formula ϕ is \mathcal{M}, S -valid iff for all $w \in W$, $w \models_S \phi$.

Definition 2. Let \mathcal{F} be a class of frames, S a semantics, a formula ϕ is \mathcal{F}, S -valid iff for all models \mathcal{M} based on a frame $F \in \mathcal{F}$, ϕ is \mathcal{M}, S -valid.

The following facts are well-known:

Fact 1 (Model necessitation for KS). *Let $\mathcal{M} = \langle W, R, V \rangle$ be a Kripke model, if ϕ is \mathcal{M},KS -valid, then $K\phi$ is \mathcal{M},KS -valid too.*

Fact 2 (Frame necessitation for KS). *Let \mathcal{F} be a class of frames, if ϕ is \mathcal{F},KS -valid, then $K\phi$ is \mathcal{F},KS -valid too.*

Of course, these facts hold as special cases for S5-models and S5-frames. However, there is major difference here between KS and CS. Frame necessitation holds for both KS and CS:

Fact 3 (Frame necessitation for CS). *Let \mathcal{F} be a class of frames, if ϕ is \mathcal{F},CS -valid, then $K\phi$ is \mathcal{F},CS -valid too.*

By contrast, as we just saw, although (ME) is valid over the model of Figure 2, the formula (KME) is not. This suffices to show that model necessitation fails for CS, namely:

Fact 4 (Failure of model necessitation for CS). *There exists a Kripke model \mathcal{M} and a formula ϕ such that ϕ is \mathcal{M},CS -valid and $K\phi$ is not \mathcal{M},CS -valid.*

It would be possible to recover model necessitation using a stronger notion of validity: following Rabinowicz and Segerberg (1994), a formula may be called *strongly valid* on a model according to CS if and only if it is true at all pairs (w, w') in the model. But note that this notion of validity is more like a model-theoretic artefact. CS is about what is true at one world. Pairs of worlds are just a notational device meant to implement the indexical character of knowledge. In this respect, the natural notion of validity on a model is just truth according to CS at all worlds (that is truth at all pairs (w, w)).

To summarize, margin for error models do validate (ME) – again, this is as it should be – but they do not validate (KME). This is because model necessitation, a familiar feature of standard Kripke semantics, fails for our centered semantics. The failure of the semantic paradox in CS is thus reflected at the syntactic level by the failure of one of the principles needed to derive the paradox. Is this a good thing? Do we have good reasons to reject (KME) while accepting (ME)? Or to put the question in the reverse direction: should the acceptance of a principle like (ME) commit us to the acceptance of a principle like (KME)?

We don't think it should be so. To use a distinction common in cognitive science, while a principle like (ME) expresses a constraint on the acquisition of knowledge of a subject whose discriminative capacities are limited, this constraint may very well be operative at the subpersonal rather than at the personal level. By contrast, a principle like (KME) requires that the subject be aware not only that her knowledge is constrained by a margin of error, but even be aware of what the value of the margin is. There is a distinction to make, in this regard, between a principle like (KK) and a principle like (KME): although both principles are iterative principles, the latter requires more than the former, since it requires that the agent be aware of the structural constraints that impinge on her knowledge. A principle like (KK), when considered in a Cartesian perspective, can be taken to reflect a notion of inner sense: whenever the subject is in a state of knowledge, then she can *thereby* know that she is in a state of knowledge. But there is no reason why a rational subject of this kind, when her knowledge is systematically constrained by a margin of error, should *thereby* know the value of the margin, even if we assume her to know that her

knowledge is imprecise. To suppose otherwise would be to suppose a capacity of self-knowledge even more powerful than Cartesian luminosity (as expressed by KK).

To be sure, Williamson himself never presents a principle like (KME) as a principle that ought to hold systematically. In his statement of the syntactic version of the paradox, Williamson simply considers that the subject “can reflect on the limitations of his eyesight and ability to judge heights”. But as emphasized in Dokic & Egré (2004), the source of this knowledge is not discussed by Williamson. We certainly agree with Williamson that the subject can reflect on his eyesight and ability to judge heights, and more generally that the subject can come to know her margin of error, but then such a knowledge is most likely acquired empirically, or simply revealed to the subject, rather than acquired a priori by some kind of inner sense. By denying the validity of (KME) over the model of Figure 2, we therefore deny that the subject can have immediate knowledge of her margin of error when (ME) holds systematically. More precisely, we saw that for every world in the model of Figure 2, there is an instance of (KME) that the subject fails to know: consequently, the subject does not know her margin of error (even though she may know particular instances of (KME): for instance, at 5, she knows $K(K\rightarrow 10 \rightarrow \neg 11)$, but that is far from sufficient, and then she in fact knows more, namely $KK\rightarrow 10$, and $K\neg 11$).

That being said, the question remains open how one should represent the knowledge of a subject who knows her margin of error systematically. To clarify this issue, we shall first present an equivalent version of (ME), which is such that it is known everywhere in the model. As we will see, however, this modified version of (ME), and the modified version of (KME) that it induces, are too weak to enable the subject to reason on her own limitations.

3.2 Actuality operators

A remarkable feature of the knowledge operator K in the framework of CS is its kinship with actuality operators. Actuality operators are operators like “now” or “actually”, which allow to preserve the reference to the time of utterance within the scope of context-shifting operators such as future or past tense operators.⁸ From a model-theoretic point of view, actuality operators can be seen as “resetting” the context of evaluation to the context of utterance. One way to implement this idea is precisely within a two-dimensional framework (Kamp 1971, Segerberg 1973). In Kamp’s temporal semantics for “now”, sentences are evaluated relative to ordered pairs (t, t') , where t represents the utterance time, and t' the event time. An atom p is true relative to (t, t') if p is true at the event time t' . Likewise, $(t, t') \models G\phi$ if and only if ϕ is true at all times later than the event time t' . By contrast, $(t, t') \models N\phi$ if and only if $(t, t) \models \phi$, namely precisely when ϕ is true at the time of utterance. In other words, “now” resets the event time to the utterance time.

Looking at the behavior of K in CS, we can see that it commands to reset the context of evaluation to the initial context in an analogous way: to evaluate a formula of the form KKp at world i , for instance, one first moves to worlds accessible from i in accordance with the standard semantics, to check whether Kp holds there; at that point, however, instead of moving to more remote worlds to check whether they satisfy p , one backtracks to i to ensure that Kp already holds

⁸See Edgington (1985): 561, from whom we borrow this formulation. We are indebted to J. van Benthem for the suggestive “resetting” image which follows.

there. In CS, the K operator thus behaves as a combination of a standard knowledge operator with an actuality operator: “I know ϕ ” is true if and only if ϕ is true at all *actual* epistemic possibilities. A closer comparison with Kamp’s semantics is appropriate to emphasize this point: in CS, an atomic sentence p is true at an ordered pair (w, w') if p is true relative to the second index, just as in Kamp’s logic; likewise, the boolean connectives are evaluated in exactly the same way; by contrast, $(w, w') \models_{\text{CS}} K\phi$ if and only if ϕ is true at all epistemic alternatives *to* w , and not to w' . Strictly speaking therefore, K behaves neither exactly like N , nor exactly like G , but as a combination of the two: like G and unlike N , K is a universal modality quantifying over a certain set of accessible worlds; unlike G but like N , K refers these worlds back to the first index.

Let us call $\mathcal{L}(K)$ the basic modal language with K as the only modal operator, and $\mathcal{L}(K, A)$ the same language augmented with a second modality A (for “actually”). Interpret $\mathcal{L}(K)$ according to the rules of CS; conversely, interpret $\mathcal{L}(K, A)$ according to the rules of the standard two-dimensional semantics, namely Kamp’s semantics in which, given a model $M = \langle W, R, V \rangle$: $(w, w') \models A\phi$ iff $(w, w) \models \phi$, and $(w, w') \models K\phi$ iff for every w'' such that $w'Rw''$: $(w, w'') \models \phi$. Finally, consider the function $*$ from sentences of $\mathcal{L}(K)$ to sentences of $\mathcal{L}(K, A)$ defined by: $p^* = p$, $(\neg\phi)^* = \neg\phi^*$, $(\phi \wedge \psi)^* = (\phi^* \wedge \psi^*)$, and $(K\phi)^* = AK\phi^*$. The function is a satisfaction-preserving map, as the following proposition shows:

Proposition 1. $M, (w, w') \models_{\text{CS}} \phi$ iff $M, (w, w') \models \phi^*$

The proof is by induction over the complexity of formulas. The atomic and boolean cases are immediate. For $\phi = K\psi$:

$$\begin{aligned} & M, (w, w') \models_{\text{CS}} K\psi \\ \Leftrightarrow & \text{for every } w'' \text{ such that } wRw'' : M, (w, w'') \models_{\text{CS}} \psi \\ \Leftrightarrow & \text{by induction hypothesis, for every } w'' \text{ such that } wRw'' : M, (w, w'') \models \psi^* \\ \Leftrightarrow & M, (w, w) \models K\psi^* \\ \Leftrightarrow & M, (w, w') \models AK\psi^* \\ \Leftrightarrow & M, (w, w') \models (K\psi)^*. \end{aligned}$$

As an immediate corollary, we obtain that $M, w \models_{\text{CS}} \phi$ if and only if $M, (w, w) \models \phi^*$. We thus obtain a translation from $\mathcal{L}(K)$ into $\mathcal{L}(K, A)$ which shows that the non-standard knowledge operator, as interpreted by CS, really is the composition of a standard knowledge operator with an actuality operator. This translation casts a new light on the results about CS mentioned in the previous section. The reader can check directly, for instance, that the complex operator AK satisfies both positive and negative introspection over pairs (w, w) with respect to the standard two-dimensional semantics, namely $AK\phi \rightarrow AKAK\phi$, and $\neg AK\phi \rightarrow AK\neg AK\phi$. Moreover, it satisfies Kripke’s axiom and the rule of necessitation over arbitrary couples: if ϕ is strongly model-valid (true relative to all pairs in a model), then so is $AK\phi$. Relative to diagonal pairs (w, w) , this is no longer the case: to repeat the point, every pair (w, w) of the model of Figure 2 satisfies the translation (ME)* of (ME), namely $AK\neg(p_i \rightarrow \neg p_{i-1} \wedge \neg p_i \wedge \neg p_{i+1})$, but not the translation (KME)* of (KME), that is $AK[AK\neg p_i \rightarrow (\neg p_{i-1} \wedge \neg p_i \wedge \neg p_{i+1})]$.

3.3 Margins actualized

By showing the correspondence between K relative to CS and AK relative to the standard two-dimensional semantics, it may seem we did not gain much. We did gain an important insight, however, since the introduction of an actuality operator shows that the problematic principles can be given a more articulated formulation without giving rise to paradox. In this section, we present another formulation of the margin principle using the actuality operator, which we call (MEA): like (ME)*, (MEA) is logically equivalent to (ME) with respect to the standard semantics and to CS; when embedded under the operator K , however, (MEA) yields a modified version of (KME) that comes out valid in the model of Figure 2.⁹ For ease of exposition, we only consider the upward version of the margin for error principle in order to formulate (MEA):

$$(MEA) \quad K\neg p_i \rightarrow A\neg p_{i+1}$$

(MEA) says that if I know that the pen is not i inches tall, the, it is actually the case that the pen is not $i + 1$ inches tall. Since the actuality operator precedes an atomic sentence, (MEA) has the same content as upward (ME). When embedded under K , however, the principle is likely to express something different from the upward version of (KME), namely:

$$(KMEA) \quad K(K\neg p_i \rightarrow A\neg p_{i+1})$$

(KMEA) says that in all my epistemic alternatives, if I know *there* that p_i is false, then p_{i+1} is *actually* false. Note that (KMEA) is much weaker than (KME) – even though, as we just said, (MEA) and (ME) equivalent. The reason is that, with (KME), the information $\neg p_{i+1}$ holds at all our epistemic alternatives in which the consequent $K\neg p_i$ is satisfied. By contrast, in (KMEA) the information $\neg p_{i+1}$ is no longer referred to the epistemic alternatives, since the actuality operator semantically extracts it from the scope of the operator K . The phenomenon is exactly analogous to the fact that, by contrast to an anaphoric pronoun like “he”, an indexical pronoun like “I” cannot be bound by the attitude-holder when embedded under an attitude operator: while “he” in “John thinks he is right” can refer back to “John”, “I” in “John thinks I am right” can only refer to the speaker and not to John.

As can be expected, using (KMEA), it is impossible to fall into the kind of soritic induction obtained by Williamson when (KME) is used instead. To be sure, let us consider what we get from $K\neg p_i$:

$$\begin{aligned} &K\neg p_i, \text{ hypothesis} \\ &KK\neg p_i, \text{ from (KK)} \\ &K(K\neg p_i \rightarrow A\neg p_{i+1}), \text{ by (KMEA)} \\ &KK\neg p_i \rightarrow KA\neg p_{i+1}, \text{ from (K)} \\ &KA\neg p_{i+1}, \text{ by (MP)} \end{aligned}$$

We derive $KA\neg p_{i+1}$ instead of $K\neg p_{i+1}$. But $KA\neg p_{i+1}$ does not express any new real knowledge, for it says that each of my epistemic alternatives is such that at the actual world p_{i+1} is false,

⁹To complete the picture, one should also consider the formulation of upward (ME) in which both antecedent and consequent are prefixed with an actuality operator, namely: $AK\neg p_i \rightarrow A\neg p_{i+1}$, which we may call (AMEA). We only focus on (MEA), however, because (MEA) and (AMEA), and their respective K -prefixed versions, are equivalent in the framework of CS.

a case of vacuous quantification. $KA\neg p_{i+1}$ does *not* say that $\neg p_{i+1}$ is true in all my epistemic alternatives, hence, it does not say that I know that $\neg p_{i+1}$.¹⁰ Formally, one cannot conclude $K\neg p_{i+1}$ from $KA\neg p_{i+1}$ (our model in Figure 2 would be a case in point).

If we now assume CS as background semantics for the richer language $\mathcal{L}(A, K)$ – in order to validate (KK) – with the additional rule that $M, (w, w') \models_{\text{CS}} A\phi$ if and only if $M, (w, w) \models_{\text{CS}} \phi$, we can see that like (KK) and (K), (MEA) and (KMEA) are both CS-valid over the model of Figure 2. The reader should compare the situation with the counterexample given earlier: we do have now that $4 \models_{\text{CS}} K(K\neg p_2 \rightarrow A\neg p_3)$, because in particular $(4, 3) \models_{\text{CS}} A\neg p_3$ (where previously we had: $(4, 3) \not\models_{\text{CS}} \neg p_3$).

With (KMEA), we are thus able to regain the idea that a rational subject can know that her knowledge is subject to a margin of error. This kind of self-knowledge does not enable the subject to increase her perceptual knowledge, however. In CS, (KMEA) and (MEA) turn out to be equivalent: both the interpretation of K in the antecedent of (MEA) and the interpretation of A in the consequent make it so that, when knowledge is iterated, the evaluation remains anchored to the actual situation. Likewise, if we consider the conclusion of the above derivation, $KA\neg p_{i+1}$ is not CS-equivalent to $K\neg p_{i+1}$, but is in fact CS-equivalent to $A\neg p_{i+1}$: within the scope of K , the actuality operator blocks the shift to further alternatives. The subject’s knowledge of her margin of error no longer gives rise to paradox. While the introduction of actuality operators is fairly drastic, the resulting set of axioms is nevertheless interesting on one aspect: arguably, it gives a model in which the knowledge of one’s margin of error is automatically satisfied, without allowing the subject to gain knowledge simply by reflecting on his perceptual capacities.

4 Comparisons

In order to close this paper, some brief comparisons will be in order, both to evaluate the plausibility and the limits of our solution to Williamson’s paradox, and also to connect the present treatment with other contextualist approaches to the phenomenon of vagueness more generally.

4.1 Varieties of knowledge

By reformulating (ME) and (KME) into (MEA) and (KMEA), we found a way to connect one plausible reformulation of Williamson’s syntactic principles with one plausible background semantics for knowledge, namely centered semantics. The good thing about this approach is that the subject can be assumed to know even the exact size of her margin of error; the limitation is that our solution may seem to restrict too much what the subject can know on the basis of this second-order knowledge. Indeed, from her knowledge that the stick is not of size 0, and the fact that she knows her margin for error, the subject can not infer that the stick is of size 1. The kind

¹⁰As pointed out by Rabinowicz and Segerberg (1994), $\phi \rightarrow KA\phi$ is weakly valid and $A\phi \rightarrow KA\phi$ strongly valid more generally. For that reason, $KA\phi$ should not be read as “I know that ϕ holds actually” (from which, in natural language, we would normally infer: “I know that ϕ ”). Rather, when true, $KA\phi$ means that the epistemic alternatives of the agent are alternatives to a situation in which ϕ holds.

of higher-order knowledge afforded by a principle like (KMEA) does not allow the subject to learn more than she did from her first-order knowledge.

In Dokic & Égré (2004), the principle (KME) is also pointed to as the suspect principle at the origin of the syntactic version of Williamson’s paradox. However, instead of using actuality operators, it is argued that Williamson failed to distinguish between kinds or methods of knowledge in his assumptions. In Dokic & Egré’s approach, Williamson’s principles are reformulated in terms of two knowledge operators, one for perceptual knowledge K_π , and the other for reflective knowledge K , with some bridge axioms to connect the two. From the assumption that $K_\pi \neg p_i$, it is possible to infer $K K_\pi \neg p_i$, and to reach the conclusion $K \neg p_{i+1}$. From that, however, it does not follow that $K_\pi \neg p_{i+1}$, as the account predicts that something can be known reflectively, without being known perceptually. Likewise, on the account presented in the previous section, from the assumption $K \neg p_i$, we reach the conclusion $K A \neg p_{i+1}$, without inferring $K \neg p_{i+1}$.

Technically, the two substitutes to (KME) bear some similarity, since both of them bring several operators into play (two knowledge operators in the one case, a single knowledge operator versus the combination of that operator with an actuality operator in the other), and both of them replace the problematic inference from $K \neg p_i$ to $K \neg p_{i+1}$ by an analogous rule involving distinct operators. Conceptually, there is a significant difference, however, since in the framework of Dokic & Egré the transition from $K_\pi \neg p_i$ to $K \neg p_{i+1}$ means that, even though it is not known perceptually that the stick is not of size $i + 1$, this can be known at least reflectively. Thus, by reflecting upon her perceptual limitations, the subject can at least make an inference about the size of the stick. In the present approach, any increase in perceptual knowledge is blocked in the same way, but arguably $K A \neg p_{i+1}$ denies the subject any kind of inferential knowledge about the stick, since in CS it is equivalent to $A \neg p_{i+1}$. To avoid this trivialization, one would need to modify the interaction of the operators A and K , in such a way that $K A \neg p_{i+1}$ can still express knowledge, without thereby collapsing into $K \neg p_{i+1}$.

The semantics proposed by Rabinowicz and Segerberg (1994), in which the accessibility relation for K is defined directly over pairs of worlds, addresses precisely this problem and it would be interesting to see how to adapt it to the case of CS. If a modified account were successfully carried out, however, the outcome would still be that K and $K A$ do not express exactly the same kind of knowledge. The approach of Dokic & Egré and the present one are perfectly consistent with each other, consequently: indeed, what we initially put into question in section 3.1 is the idea that a subject whose knowledge is constrained by a margin of error principle can thereby know her margin of error. More specifically, by reformulating (ME) and (KME) into (MEA) and (KMEA) in section 3.3, we gave a model in which the reflective knowledge of the margin is automatically satisfied, but without allowing the subject to extend her knowledge. This model is consistent with the view that, without an additional empirical input, reflection upon one’s perceptual capacities should not allow one to extend one’s knowledge. But if, finally, the subject is to *learn*, from experience or from some oracle, that her knowledge obeys a specific margin of error, then we agree that the subject might be able to use this additional knowledge and extend her initial perceptual knowledge.

In that case too, however, we remain in agreement with the conclusions of Dokic & Egré: a model in which the agent learns her margin should be a model in which the sources and varieties of knowledge are explicitly distinguished. Suppose the subject were to learn from some infallible

oracle that her initial knowledge is subject to a margin of error, revealing to her the size of the margin: then obviously this second-order knowledge does not stand on the same footing as her initial knowledge, since it is acquired from an external source and not subject to any error by definition. But even in the case in which the subject learns from experience and careful empirical reflection upon her knowledge that her knowledge is subject to a margin of error, there are good reasons to doubt that this second-order knowledge is of the same kind as her initial knowledge. This new kind of knowledge may also be subject to error and approximations, but the method by which this knowledge is acquired is most likely not the same as that which underlies her first-order knowledge.

Our account may therefore be summarized by the following list of options: if the subject's knowledge obeys (ME), (KK) and (K), then the subject may simply fail to know what her margin is, as is reflected when CS is adopted as background semantics; but if the subject is to know her margin a priori, as happens when (ME) and (KME) are rephrased as (MEA) and (KMEA), then the subject's knowledge of her margin should not give her more information than is given to her by her first-order knowledge; finally, if a subject whose knowledge is described by (ME), (KK) and (K) is to learn the information that her knowledge obeys (ME), then we agree that this information may take her initial knowledge further, but our claim is that this knowledge is not homogeneous with the initial knowledge to which (ME) applies.

4.2 Contextualism and the sorites

The second question we want to address concerns semantic relativity, and the link of the present account to contextualist treatments of vagueness. By semantic relativity, we refer to the fact that the validity or invalidity of principles like (KK) and (KME) depends on a choice between distinct semantic characterizations of knowledge. Williamson's paradox shows the incompatibility of the introspection principle (KK) when margin for error principles are assumed to hold unrestrictedly for knowledge. In particular, we saw that Williamson's margin for error semantics, based on the standard semantics for knowledge, invalidates the axiom (KK). By contrast, our centered semantics makes (KK) valid, but we saw that it does not validate all margin for error principles: thus, the model of Figure 2 validates (ME), but invalidates (KME). Which semantics is the right semantics? The first thing to say is that both semantic characterizations of knowledge may equally fail to be right. Indeed, both treatments of knowledge rest on strong idealizations and consequently both are likely to yield an overly simplified description of the situations of inexact knowledge that we discussed.¹¹ But the question then should be: which of the two accounts of knowledge seems more likely to be right? To our minds, centered semantics gives a richer model of the context-dependence of margin for error principles for the kind of scenarios we discussed here. To a large extent, the soritical progression induced by the iteration of K modalities is an artefact of the standard Kripke-Hintikka semantics. To repeat the point of the previous section: knowing that ϕ , in centered semantics, means that ϕ holds at all *actual* epistemic alternatives. This anchoring to the initial context, we believe, gives a more realistic account of the situated

¹¹One such idealization concerns the use of fixed margin of errors, irrespective of the quality or intensity of the stimulus in our model.

character of knowledge and belief.

By way of conclusion, and to give another perspective on the contextualist epistemic semantics proposed here, we may point out an analogy between our treatment of Williamson’s paradox and Kamp’s influential treatment of the sorites in Kamp (1981).¹² In this paper, Kamp presents two thought experiments, in which a subject faces a screen divided in small squares, together displaying a gradual transition from green to yellow. In the first thought experiment, the screen is completely visible and the subject is asked of each successive square whether it is green. In the second thought experiment, the screen is covered, and only pairwise adjacent squares are uncovered successively. Kamp supposes that the judgments will vary from one experiment to the other. In the first experiment, in particular, Kamp notes that the extreme squares can serve as “anchor points” to establish a comparison with the focal square, while only the adjacent square can do so in the second experiment. He is led to think for that reason that the judgments will shift more quickly from “Green” to “Non-Green” or “Uncertain” in the first experiment than in the second experiment. Kamp’s semantics is thus an attempt to make each instance of the tolerance principle $\forall n(P(n) \rightarrow P(n + 1))$ true, while making the universal claim itself false. His idea is that each instance holds locally, as it were, but not globally.

Interestingly, we get the same kind of effect in CS when we compare (ME) to (KME): in the model of Figure 2, we saw that the conditional $K\neg p_i \rightarrow \neg p_{i+1}$ is universally true; however, its universal generalization $K(K\neg p_i \rightarrow \neg p_{i+1})$, which in combination to (K) and (KK) plays exactly the role of an induction principle, is not valid. If we ponder on the reasons of this failure, we can observe that they depend in an analogous way on a shift in “anchor point” (the reference point in the terminology we used here). Thus, $K\neg p_2 \rightarrow \neg p_3$ holds at 4 because it holds at (4, 4). By contrast, $K(K\neg p_2 \rightarrow \neg p_3)$ does not hold at 4, because its truth now depends on the pair (4, 3), which makes $K\neg p_2 \rightarrow \neg p_3$ false. We thus get the same kind of contextual effect which Kamp imagines in his thought experiments, albeit on a smaller scale: when the perspective point and the reference point coincide, (ME) is valid; when the reference point deviates sufficiently from the perspective point, (ME) can be invalidated.

This analogy, of course, is not meant to undermine the profound conceptual differences between Kamp’s view of the sorites (which preserves a local form of the tolerance principle) and the view of the epistemicists (who simply deny that all instances of the tolerance principle are true). However, our aim here was not to present an overall solution to sorites paradoxes in general. Our aim was rather to untie the knot of one specific sorites, namely the epistemic sorites which results within the metatheory of the epistemicists themselves when margin for error principles – which are used to account for the origin of vagueness – are assumed in combination with introspection principles. Despite this, we are inclined to think that the ingredients of our own contextualist solution to Williamson’s paradox enforce the general paradigm of contextualist approaches to vagueness. It would be interesting, in particular, to examine whether the kind of multi-dimensional semantics we used for an epistemic language can be transposed to a regular first-order language in order to handle the sorites paradox more generally, a problem we leave for further work.

¹²Other contextualist approaches include in particular Soames (1999), Graff (2000) and Shapiro (2006).

Acknowledgments

We wish to thank the Editors, Mak Kölbel and Manuel Garcia-Carpintero, for their invitation and for encouraging us to investigate the contextualist dimension of our epistemic semantics. At the time we came up with the first version of Centered Semantics, our concern was primarily to have a modular semantics for knowledge and a plausible model of knowledge iterations; we became aware of the link with indexicality issues only secondarily. Special thanks are due to P. Schlenker for pointing out the link with Kaplanian themes, to J. van Benthem for directing our attention to Kamp's temporal logic, and to W. Rabinowicz for referring us to his work with K. Segerberg. We also thank M. Cozic, J. Dokic, and an anonymous referee for helpful comments, as well as audiences in Lisbon (ENFA3), Prague (Prague Colloquium on *Reasoning about Vagueness and Uncertainty*) and Amsterdam (PALMYR 4).

References

- [1] Bonnay D. & Égré P. (2006), A Non-Standard Semantics for Inexact Knowledge with Introspection, *Proceedings of the ESSLLI 2006 Workshop on Rationality and Knowledge*, S. Artemov & R. Parikh (eds).
- [2] Dokic J. & Égré P. (2004), Margin for Error and the Transparency of Knowledge, Technical Report, Electronic archives of the Jean-Nicod Institute, <http://jeannicod.ccsd.cnrs.fr/ijn.00000479>.
- [3] Edgington D. (1985), The Paradox of Knowability, *Mind*, 94, 557-68.
- [4] Égré P. (forthcoming), Reliability, Margin for Error and Self-Knowledge, forthcoming in D. Pritchard & V. Hendricks (eds.), *New Waves in Epistemology*, Ashgate.
- [5] Fara D. Graff (2000), Shifting Sands: An Interest-Relative Theory of Vagueness, *Philosophical Topics*, vol. 28(1): 45-81. Originally published under the name “Delia Graff”.
- [6] Fara D. Graff (2002), An Anti-Epistemicist Consequence of Margin for Error Semantics for Knowledge, *Philosophy and Phenomenological Research*, 64, 127-142. Originally published under the name “Delia Graff”.
- [7] Gomez-Torrente M. (1997), Two Problems for an Epistemicist View of Vagueness, *Philosophical Issues*, 8, *Truth*, 237-245.
- [8] J. Halpern (2004), Intransitivity and Vagueness, *Proceedings of the Ninth International Conference on Principles of Knowledge Representation and Reasoning (KR 2004)*, 121-129.
- [9] Kamp H. (1971), Formal Properties of “Now”, *Theoria* 37, 227-273.
- [10] Kamp H. (1981), The Paradox of the Heap, in U. Mönnich (ed.), *Aspects of Philosophical Logic*, 225-277.
- [11] Kaplan D. (1989), Demonstratives, in J. Almog & al. (eds), *Themes from Kaplan*, Oxford.
- [12] Kölbel M. (forthcoming), Vagueness as Semantic, forthcoming in R. Dietz & S. Moruzzi (eds), *Cuts and Clouds: Essays in the Nature and Logic of Vagueness*, Oxford University Press.
- [13] Rabinowicz W. & Segerberg K. (1994), Actual Truth, Possible Knowledge, *Topoi* 13, 101-15.
- [14] Recanati F. (2000), *Oratio Recta, Oratio Obliqua*, MIT Press.
- [15] Segerberg K. (1973), Two-Dimensional Modal Logic, *Journal of Philosophical Logic*, 2, 77-96.
- [16] Shapiro S. (2006), *Vagueness in Context*, Oxford University Press, Oxford / New York.

- [17] Soames S. (1999), *Understanding Truth*, Oxford University Press, Oxford / New York.
- [18] Sorensen R. (2001), *Vagueness and Contradiction*, Oxford, Clarendon Press.
- [19] Williamson T. (1992), Inexact Knowledge, *Mind*, 101, 217-42.
- [20] Williamson T. (1994), *Vagueness*, Routledge.
- [21] Williamson T. (2000), *Knowledge and its Limits*, Oxford University Press.