Vagueness and Introspection

Denis Bonnay, Paul Egré

▶ **To cite this version:**

HAL Id: ijn_00141415

http://jeannicod.ccsd.cnrs.fr/ijn_00141415

Submitted on 12 Apr 2007

Vagueness and Introspection*

Denis Bonnay   Paul Égré

Abstract

We compare three strategies to model the notion of vague knowledge in epistemic logic. Williamson's margin for error semantics typically uses non-transitive Kripke structures, but invalidates the principle of positive introspection. On the contrary, Halpern's two-dimensional semantics preserves the introspection principle, but using more complex uncertainty relations that are transitive. We present a modification of the standard epistemic semantics, which validates introspection over one-dimensional non-transitive structures, and study its correspondence with Halpern's approach. While the semantics can be seen as the diagonalization of an explicit two-dimensional semantics, it affords a more intuitive representation of the uncertainty characteristic of vague knowledge. We examine the implications of the semantics concerning higher-order vagueness and the status of the non-transitivity of perceptual indiscriminability. We respond to a potential objection against our approach by giving a dynamic model of the way subjects with inexact knowledge make successive approximations of their margin of error.

1   Intransitivity and introspection

One central and debated aspect of the notion of inexact knowledge concerns the non-transitivity of the relation of indiscriminability and how it should be represented. On the epistemic account of vagueness put forward by Williamson, the intransitivity of the relation of indiscriminability is presented as the main source for vagueness ([12]: 237). In [11] and in the Appendix to [12], Williamson formulates a *fixed margin for error* semantics for propositional modal logic in which the relation of epistemic uncertainty,

---

based on a metric between worlds, is thus reflexive and symmetric, but non-transitive and non-euclidian.[1] An important consequence of Williamson's semantics is that it invalidates the principles of positive introspection (if I know p, then I know that I know p) as well as negative introspection (if I don't know p, then I know that I don't know p).

In an earlier paper [1], we argued against Williamson that models of inexact knowledge that preserve the introspection principles can sometimes be desirable, and we presented a non-standard epistemic semantics for the notion of inexact knowledge, in which non-transitive and non-euclidian Kripke models can nevertheless validate positive as well as negative introspection. In [5], Halpern also argued against Williamson that an adequate model of vague knowledge need not invalidate the introspection principles, but following a different route. Instead of taking intransitivity as a primitive, and proving that the introspection principles can be preserved for a logic with one epistemic operator, as we did in [1], Halpern proposes a bimodal account of inexact knowledge that preserves the introspection principles, and he shows that there is a way to derive intransitivity. For Halpern, the intransitivity of vague knowledge is more characteristic of our reports on what we perceive than about our actual perception.

Despite these differences, one can establish a precise correspondence between Halpern's semantics and the semantics presented in [1]. The object of this paper is to spell out the details of this correspondence, and thus to compare two strategies in order to keep together introspection and non-transitivity. Like Halpern, but contra Williamson, we think it does make sense to preserve the introspection principles within a logic of inexact knowledge; unlike Halpern, but in agreement with Williamson, we are ready to see non-transitivity as a property of perceptual knowledge proper. The divergence with Halpern's view is more conceptual than technical, however, since we will see that we understand "perceptual" in a broader sense than Halpern, in a way that encompasses what he calls "reports about perception". More fundamentally, however, our approach and Halpern's both rest on the idea that the non-transitivity of phenomenal indiscriminability can be made transitive by reference to a particular context, but in our approach this contextual parameter remains implicit, thereby affording a more intuitive characterization of the uncertainty associated to inexact knowledge.

The paper is organized as follows: in Section 2, we introduce Centered Semantics, a non-standard semantics for modal epistemic logic, and review how it can be combined with Williamson's margin semantics in a way that

---

[1] A relation R is transitive if xRy and yRz imply xRz for every x, y and z. A relation is euclidian if xRy and xRz imply yRz for every x, y, z.

preserves the introspection principles. In Section 3, we examine the correspondence with Halpern's semantics. In Section 4, finally, we examine a possible objection against the semantics introduced in Section 2, namely the fact that the semantics invalidates a principle of knowledge of one's margin of error put forward by Williamson in his attack against the principle of positive introspection. We respond to the objection by presenting a dynamic model of the way in which subjects with inexact knowledge make successive approximations of their own margin of error.

2   Centered Semantics

Consider a discrete series of pens linearly ordered by size, such that all and only pens that are less or equal to 4 cm fit in a certain box. A subject sees the pens and the box at a certain distance and is asked which pens will fit in the box. We make the supposition that from where she is, the subject cannot perceptually discriminate between pairwise adjacent pens, namely between pens whose size differs by less than 1 cm. However, the subject is able to discriminate between non-adjacent pens.[2] For instance, when looking at the pen of size 2, the subject cannot discriminate it from the pen of size 1, nor from the pen of size 3, but she can discriminate it from a pen of size 4. Furthermore, we make the idealized assumption that the inability of the subject to detect differences is constant throughout the series.

The scenario may be represented by means of the following linear Kripke model, in which p represents the objective property of fitting in the box, with worlds indexed by sizes. The important fact about the model, reflected in the accessibility relation between worlds, is that the model is reflexive and symmetric, but non-transitive (and non-euclidian). Letting R stand for the relation of perceptual indicriminability, this represents the fact that the subject does not discriminate any object from itself, nor any two adjacent items in the series, but can indeed discriminate between any two non-adjacent items.
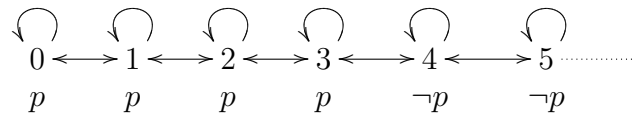

Figure 1: A discrete margin model

---

[2]The size unit is not relevant for the discussion, and the reader may replace centimeters by millimeters if it helps make the scenario more plausible. Likewise, sizes may be translated for more plausibility (translating 0 to 18, 1 to 19 and so on).

We use the language of propositional modal logic, where $\Box$ is interpreted as a knowledge operator: $\Box\phi$ stands for: "the subject knows that ϕ". We consider the usual semantics for propositional modal logic, in which given a Kripke model $\langle W, R, V \rangle$, and a world $w \in W$, $w \models \Box\phi$ iff for every w' such that wRw', $w' \models \phi$. Relative to this model, $n \models \Box p$ means that when looking at a pen of size n, the subject knows that it fits in the box. In the present case, this means that when looking at a given object, the subject knows that it has some property if and only if every object that is indiscriminable from it has that property. Alternatively, the above model can be seen as a particular case of what Williamson calls a fixed margin model. A *fixed-margin* model is a model $\langle W, d, \alpha, V \rangle$, where d is a metric over W, and α a real valued margin for error parameter, such that $w \models \Box\phi$ iff for every w' such that $d(w, w') \leq \alpha$, $w' \models \phi$. That is, at a given world, the subject knows that some proposition ϕ holds if ϕ holds at every world included within the margin α from w. The above model is a discrete margin model, such that $W = \mathbb{N}$ and $\alpha = 1$.

In the above model, for instance, $2 \models \Box p$, and $3 \models \neg\Box p \wedge \neg\Box\neg p$: thus, the subject knows that an object of size 2 will fit in the box, and does not know whether an object of size 3 fits in the box. Crucially, however, $2 \models \neg\Box\Box p$, that is the subject doesn't know that he knows that the pen fits in the box, since 4 is a $\neg p$-world accessible in two steps from 2. For Williamson, this result is a welcome prediction of the model and semantics, since iterations of knowledge operators are seen by Williamson as a "process of gradual erosion" in the case of vague knowledge ([12]: 228). Indeed, each iteration of knowledge is seen as a step by which a margin of error is removed. According to Williamson, the subject knows that he knows that p only if his knowledge is "safely safe", namely if the epistemic context of the subject is at least two steps away from the boundary between p and $\neg p$.[3]

However, one may argue that, looking at a pen of size 2, my knowing that I know that it will fit in the box supervenes only on my knowing whether it fits in the box, and not on epistemic alternatives that are further away. One important motivation to suppose so concerns higher-order iterations of knowledge: for instance, the standard semantics makes the prediction that at 0 in the above model, it holds that $\Box\Box\Box p$, and yet that it is not the case that $\Box\Box\Box\Box p$. However, it is hard to make sense of such fine-grained distinctions between levels of knowledge: indeed, if the subject knows that she knows that she knows that the pen fits in the box, how could she fail to know that she knows that she knows that she knows?

In [1], we formulated an alternative semantics (CS, for Centered Semantics), in which the epistemic alternatives relevant for iterated modalities re-

---

[3]See [14], p. 123 on the topological understanding of the notion of safety.

main the worlds accessible in one transition from the world of evaluation. In other words, every fact concerning the knowledge of the agent should be decided solely on the basis of worlds that are not distinguishable from that world, without having to move further along the accessibility relation. Given a model $M = \langle W, R, V \rangle$, we first define the notion of satisfaction for couples of worlds, and extract the definition of satisfaction for single worlds:

Definition 1. *CS-satisfaction for couples of worlds:*

(i) $M, (w, w') \vDash_{CS} p$ iff $w' \in V(p)$.

(ii) $M, (w, w') \vDash_{CS} \neg\phi$ iff $M, (w, w') \nvDash_{CS} \phi$.

(iii) $M, (w, w') \vDash_{CS} (\phi \wedge \psi)$ iff $M, (w, w') \vDash_{CS} \phi$ and $M, (w, w') \vDash_{CS} \psi$.

(iv) $M, (w, w') \vDash_{CS} \Box\phi$ iff for all w'' such that wRw'', $M, (w, w'') \vDash_{CS} \phi$.

Definition 2. $M, w \vDash_{CS} \phi$ *iff* $M, (w, w) \vDash_{CS} \phi$

The use of double-indexing allows us to represent both the perspective of the agent (through the first index, which we may call the *perspective* point), and also the information relevant relative to the agent's perspective (through the second index, which bears the atomic information, and which we may call the *reference* point).[4] Satisfaction with respect to single worlds is defined by diagonalization, namely when the perspective point and the reference point coincide. Thus clause (iv) of Definition 1 and Definition 2 together account for the "centered" feature of the semantics, for they entail that for every w and w': $M, (w, w') \vDash_{CS} \Box\phi$ iff $M, (w, w) \vDash_{CS} \Box\phi$ iff $M, w \vDash_{CS} \Box\phi$. Thus, instead of looking at worlds that are two steps away to check whether $\Box\Box\phi$ is satisfied, one backtracks to the actual world to see whether $\Box\phi$ already holds there. In the previous model, it can be checked that $2 \vDash_{CS} \Box p$, and likewise $3 \vDash_{CS} \neg\Box p \wedge \neg\Box\neg p$. However, $2 \vDash_{CS} \Box\Box p$ and $3 \vDash_{CS} \Box\neg\Box p$. In [1], we proved that the normal logic K45 is indeed sound and complete with respect to CS.[5] Furthermore, one can formulate a centered version of Williamson's fixed-margin semantics, which we call CMS, for which the logic S5 is sound and complete (we shall not repeat the proofs here, but refer to [1] for technical details).

---

[4]See [2] for details on the relation of CS to other double-indexing frameworks, in particular Rabinowicz & Segerberg's in [8]. The terminology of *perspective* vs *reference* points is from Rabinowicz & Segerberg.

[5]Axiom K is $\Box(p \rightarrow q) \rightarrow (\Box p \rightarrow \Box q)$, 4 is $\Box p \rightarrow \Box\Box p$ (positive introspection) and 5 is $\neg\Box p \rightarrow \Box\neg\Box p$ (negative introspection). S5 is the extension of K45 obtained by adding T, namely $\Box p \rightarrow p$.

When $\Box$ is interpreted as "the subject knows that" or "I know that", our view is that Centered Semantics gives a more plausible view of higher-order knowledge than the standard semantics. When the operator $\Box$ is interpreted as "it is clear that", however, as Williamson originally considered in his Logic of Clarity, the situation may be different. In particular, the main motivation to conceive of iterations as a process of gradual erosion is to have an account of higher-order vagueness: the idea is that clear cases of some property, like unclear instances, may in turn have borderline cases. For instance, $2 \models \Box p$, which under this interpretation means that 2 clearly fits in the box, but $2 \nvDash \Box\Box p$, which means that it isn't clear that 2 clearly fits in the box. Relative to this interpretation of the $\Box$, it can therefore be objected that CS makes room only for first-order vagueness, and not for higher-order vagueness, since "it is clear that p" systematically entails "it is clear that it is clear that p" in CS (positive introspection), and furthermore "it is not clear that p" systematically entails "it is clear that it is not clear that p" (negative introspection).

To this, however, three replies can be made: firstly, when $\Box$ is read as "I know that", as we assume, $\Box\neg\Box p$ should rather mean that I am aware of my uncertainty at the moment it first arises: by analogy to situations of "forced march" (see [6]: 173), in which I am forced to answer by "yes" or "no" (or possibly more values) for the application of one and the same predicate of each item in a soritical series, this means that I am aware of making a "jump" in my judgements when the jump occurs (see e.g. [10]).[6] In the present case, we could imagine that the choice is forced between "yes", "no" and "indeterminate". In that case, the subject will say "yes" to "does pen 0 fit in the box?", and likewise for pens 1 and 2. When looking at pen 3, however, the subject starts to hesitate: the subject then "jumps" and answers "I'm not sure" (namely "indeterminate"). But at the moment the subject makes this jump from "yes" to "indeterminate", the subject presumably is aware of making the jump: it is perfectly consistent with the scenario to imagine that the subject knows that she doesn't know whether 3 fits in the box or not.

A second element of response to the problem of higher-order vagueness concerns the fact that in [1] we show that CS is a particular case of a family of resource-sensitive semantics called TS(n) (for "token semantics with n tokens"), for which the trivialization of the iterations need not occur at the first level, but at any arbitrary level n of iterated modalities, depending on the number n of tokens available. Informally, the intuition behind To-

---

[6]A jump occurs as soon as the subject declares a difference in semantic status for two subsequent items in the series. See [6], 173 sqq.

ken Semantics is that moving along the accessibility relation has a cognitive cost, which is mirrored by the fact that a token is spent for each move in a model, and the initial number of tokens available to the agent is finite. When all tokens have been spent, just as in CS, the agent backtracks to the position reached before the last move, gets a token back, and can spend it for a new move. In this framework $w \vDash_{\text{TS}} \phi [n]$ means that ϕ holds at w when n tokens are available to the agent (we refer to the Appendix for the recursive definition). To take a concrete example on the previous model, the rules of satisfaction in Token Semantics predict that: $1 \vDash_{\text{TS}} \Box\Box p$ [2], as in the standard semantics, since all worlds reachable in two steps from 1 satisfy p. Likewise, as in the standard case, $2 \nvDash_{\text{TS}} \Box\Box p$ [2]. However, we now have that $1 \vDash_{\text{TS}} \Box\Box\Box p$ [2], since with only 2 tokens, the agent cannot visit worlds beyond 3 when her initial context is 1. It is easy to see that CS corresponds to Token Semantics with only 1 token allowed, and standard Kripke Semantics to Token Semantics with an infinite number of tokens available. More generally, $\text{TS}(n)$ and standard Kripke semantics coincide for formulas with less than n embedded modalities.

Using $\text{TS}(n)$ semantics, one can in principle account for n-order vagueness when the $\Box$ operator is interpreted as "it is clear that". But again, $\text{TS}(n)$ cannot be a logic of $n+1$-order vagueness, since in $\text{TS}(n)$ the schema $\Box^n p \rightarrow \Box^{n+1} p$ comes out as a validity (see [1] and the Appendix for technical details): thus "it is clear $(n-1$ times)... that p" does not necessarily entail "it is clear $(n$ times) that p", but "it is clear $(n$ times) that p" entails that "it is clear that it is clear $(n$ times) that p". The question, here, is whether one can plausibly conceive of higher-order vagueness without being committed to higher-order vagueness at all orders. Williamson, for instance, shows that in KTB, namely the basic Logic of Clarity, a proposition p either is precise, or has first-order vagueness but not higher-order vagueness, or has vagueness of all orders ([13]:136).[7] However, Williamson remarks that if "B is abandoned, p can have vagueness of all orders below n and precision thereafter for any $n \geq 1$" ([13]:138). Conversely, Williamson notes that in a logic stronger than KTB, like S5, there can be first-order vagueness without second-order vagueness. The relevant point here is that for every $n \geq 2$, $\text{TS}(n)$ is axiomatized by a logic weaker than K45 that fails to yield B, even when T is systematically included. Conversely, when T and B are assumed, the resulting logic, which includes the schema $\Box^n p \rightarrow \Box^{n+1} p$ (see Appendix) is then stronger than KTB. Either way, the resulting logic for $\text{TS}(n)$ remains a plausible candidate

---

[7]B is the axiom $p \rightarrow \Box\neg\Box\neg p$. KTB axiomatizes Williamson's fixed margin semantics. Williamson also presents a variable margin semantics, axiomatized by the logic KT.

for n-order vagueness.[8] More fundamentally, we do consider that higher-order vagueness running out at some finite order is plausible enough, much for the same reasons which concern the iterations of knowledge operators. In our view, instances of a property may become clearly clear instances at some point, in much the same way in which a man with 0 hair on his head is a clear cases of baldness, and a clearly clear case thereof, and so on at all levels (see [4]).

The third remark on this problem, finally, concerns the fact that the operators "it is clear that" and "I know that" should not be taken to be synonymous, even when the kind of knowledge described is inexact.[9] Indeed, taking CS as a logic of inexact knowledge does not commit us to denying higher-order vagueness, so long as the $\square$ operator is interpreted as "I know that", and not as "it is clear that". This point is central to Halpern's own approach to the representation of vagueness in epistemic logic, to which we now turn.

3   Halpern's semantics

Halpern takes a different approach to the problem of inexact knowledge, since his logic makes room for distinct syntactic representations of the operators "I know that" and "it is clear that". His logic (in the one-agent case) has two primitive operators, namely R and D, where $R\phi$ means that the agent "reports ϕ", and $D\phi$ means that "according to the agent, ϕ is definitely the case". A model, relative to this language, is a structure $\langle W, P, \sim_s, \sim_o, V \rangle$, where $W \subseteq O \times S$, where S is intended to denote a set of subjective states and O a set of objective states. The relations $\sim_o$ and $\sim_s$ both are equivalence relations over W, and V is a valuation over W. P, finally, is a subset of W, intended to denote the states that the agent considers plausible. For simplicity, we shall assume that $P = W$ here, and therefore we shall omit reference to P in the definition of satisfaction. With that simplification, the satisfaction clauses for the modal operators are the expected ones, namely $M, (w, v) \models R\phi$ iff for every (w', v') such that $(w, v) \sim_s (w', v')$, $M, (w', v') \models \phi$, and similarly for $D\phi$ with respect to $\sim_o$. As a consequence,

---

[8]Williamson [13], p. 134, notes that "intuitively, there is a strong connection between the non-transitivity of indiscriminability and higher-order vagueness". What matters in this respect is that while transitivity and euclidianness are invalid in TS(n) for $n > 1$, weaker forms of both properties are preserved for the logic, namely n-transitivity and n-euclidianity, as described in the Appendix.
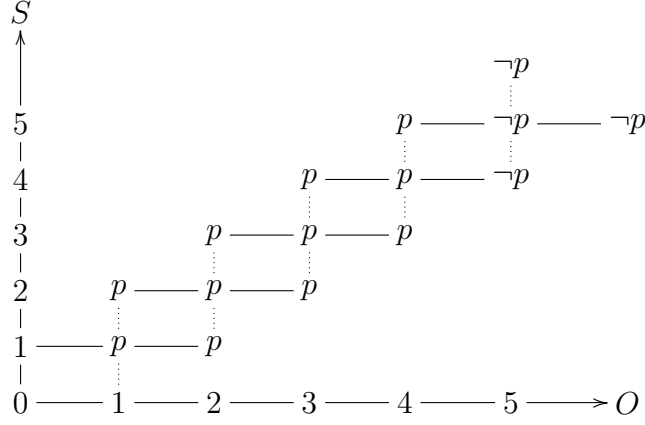
[9]See [2] for further developments on this point.

Figure 2: A layered margin model

each operator is axiomatized by the logic S5.[10] The point of Halpern's approach, however, is that although each operator separately obeys transitivity (and euclidianness), their combination DR need not (if two binary relations A and B are equivalence relations, it does not follow that their composition $A \circ B$ is transitive or euclidian).[11] Intuitively, an agent definitely reports that ϕ when his estimation is sufficiently reliable, just as in Williamson's approach. In this way, the complex operator (DR) plays exactly the role of Williamson's "clearly" operator in margin for error semantics.

To make the link concrete, let us consider a model, depicted in Figure 2, in which W is the subset of $\mathbb{N} \times \mathbb{N}$ consisting of couples (n, m) such that $|n - m| \leq 1$. Let us suppose that n is the objective size of some object, or the objective value of some parameter, and m its subjective estimate. The constraint on n and m represents the fact that the subject's estimate cannot deviate from more than 1 on the objective value, namely that the subject's margin of error is 1. Let us suppose moreover that $(n, m) \sim_o (n', m)$ iff $n = n'$, namely if they agree on their objective indices, and likewise $(n, m) \sim_s (n', m')$ iff $m = m'$, namely if they agree on their subjective indices. It is easy to verify that both relations are equivalence relations over W.

In the above figure, each cell of the partition determined by $\sim_o$ corresponds to the points connected by a vertical dotted line, and each cell of the partition determined by $\sim_s$ corresponds to the points connected by a

---

[10]In Halpern's full version of the semantics for the multi-agent case, each modality is actually a KD45 operator (where D is $\Box \neg p \to \neg \Box p$), and for each agent the corresponding D operator satisfies a weakened version of axiom **T**.

[11]For two binary relations A and B over W, $A \circ B =_{df} \{(v, w); \exists u : (v, u) \in A \wedge (u, w) \in B\}$.

horizontal straight line. Let us suppose moreover that whether a point w is a member of $V(p)$ depends only on the objective part of w. For instance suppose that $(n, m) \in V(p)$ iff $n < 5$ (as in our previous example, p may stand for "fitting in the box"). It is easily checked that $(2, 3) \models DR\ p$, but $(2, 3) \nvDash DRDR\ p$. Thus, if the size of the object is 2 and the measurement made by the agent is 3, with a threshold for $\neg p$ that is between 4 and 5, then the agent definitely reports that p, but will not iterate this judgement. By contrast, R is an S5 modality, satisfying negative and positive introspection at any point in the model.

In the previous model, Dp is equivalent to p. Conversely, if we consider only the relation of *subjective* equivalence for R, a model like the model of Figure 2 may be called a *layered margin model*, since each horizontal equivalence class (namely the classes for $\sim_s$) contains the possible objective values that are compatible with the agent's subjective parameter, and the horizontal projection of these classes onto the O-axis of the model would yield a linear structure of inexact knowledge of exactly the kind with which we started. This notion of layering can be made precise. Thus, given a Kripke model $M = \langle W, R, V \rangle$, let us call $L(M) = \langle W', R', V' \rangle$ a *layering* of M if it satisfies: $W' = \{(w, w') \in W \times W; w'Rw \lor w' = w\}$; $(w, w')R'(u, u')$ iff $w' = u'$ and $w'Ru$; and finally, $(w, w') \in V'(p)$ iff $w \in V(p)$ (note that it is the first index here, namely the objective index, which specifies the atomic information, whereas in Centered Semantics the first index, or perspective point, comes first: this explains the inversion of indices in Lemma 1 below). It can be checked that R' in $L(M)$ is necessarily transitive and euclidian for every R, and is an equivalence relation if R is reflexive. It is easy to establish that, relative to the basic modal language in which $\Box$ is the single modality:

Proposition 1. $M, w \vDash_{\mathrm{CS}} \phi \iff L(M), (w, w) \models \phi$

by proving that for all (w', w) in $L(M)$:

Lemma 1. $M, (w, w') \vDash_{\mathrm{CS}} \phi \iff L(M), (w', w) \models \phi$.

Proof. $M, (w, w') \vDash_{\mathrm{CS}} p$ iff $w' \in V(p)$ iff $(w', w) \in V'(p)$ iff $L(M), (w', w) \models p$. The boolean cases are immediate. $M, (w, w') \vDash_{\mathrm{CS}} \Box\phi$ iff for every w'' such that wRw'', $M, (w, w'') \vDash_{\mathrm{CS}} \phi$ iff, by induction hypothesis, for every w'' such that wRw'', $L(M), (w'', w) \models \phi$, iff, by definition of $L(M)$, for every (u, u') such that $(w', w)R'(u, u')$, we have $L(M), (u, u') \models \phi$, iff $L(M), (w', w) \models \Box\phi$. $\square$

Proposition 1 applies also to Williamsons's fixed margin models $M = \langle W, d, \alpha, V \rangle$ (see [4]) in which two worlds w, w' are accessible iff $d(w, w') \leq \alpha$, where d is a metric over W. A layered margin model necessarily is a model

in which the induced epistemic accessibility is an equivalence relation, since margin models are reflexive by definition of a metric between worlds. With respect to margin models, Proposition 1 shows that Halpern's operator R therefore plays exactly the role of the knowledge operator $\Box$ in the framework of centered semantics. On the one hand, these results makes clear that centered semantics really is a standard two-dimensional semantics in disguise. On the other hand, the operation of layering shows how it is possible to recover transitivity (and euclideanness) from a non-transitive (and non-euclidian) relation. In Halpern's approach, the intransitivity characteristic of qualitative comparison is simulated by means of two operators.

According to Halpern, if one conceives of epistemic accessibility relations as indistinguishability relations, "there is a strong intuition that the indistinguishability relation should be transitive, as should the relation of equivalence on preferences" ([5]: 2). Halpern's arguments in favour of this intuition converge with the arguments given independently by Fara in [3] and Raffman in [9], in favour of the idea that perceptual indiscriminability is transitive, despite appearances to the contrary. For Fara, in particular, the apparent non-transitivity of "looking the same as" is due to a surreptitious shift of the context with regard to which judgments of resemblance are made. Raffman reaches the same conclusion as Fara, by pointing to the fact that in soritical series, sameness is context-dependent, and that one of the same color patch #4, for instance, may upon reflection look different in two different acts of comparison, even though it looks the same as #3 in one context, and the same as #5 in a different context.

The correspondence between Halpern's approach and ours suggests that we can perfectly agree with Halpern, Fara or Raffman on the idea that phenomenal indiscriminability is transitive *once it is explictly contextualized*. Indeed, in a layered margin model such as the model of Figure 2, the relation of subjective indistinguishability is between ordered pairs with the same objective index. In the non-transitive margin model of Figure 1, by contrast, the relation of indistinguishability is not relativized in this way. However, when epistemic sentences are evaluated with respect to Centered Semantics over this model, what obtains is in fact an *implicit relativization* of exactly the same kind: to say that 1 is indistinguishable from 0 and from 2 (in CS) in fact means that 1 is indistinguishable from 0 and 2 *when 1 is taken as perspective point*.[12]

---

[12]In [2], we made this contextual effect explicit in a different way: instead of viewing Centered Semantics as an implicit transformation of one-dimensional model into a product model, as shown in Proposition 1, we prove that Centered Semantics can be seen as an implicit enrichment of the underlying epistemic language by means of an actuality operator. More precisely, what we show is that "knowing", relative to CS, means the

Despite this, one could still ask which of the two indistinguishability relations should be considered as more primitive in order to describe knowledge: should we take as primitive the transitive relation of indistinguishability of the layered model? or rather, should it be the non-transitive relation of the original margin model? Here Proposition 1 can be interpreted in two opposite ways. Like Halpern, we may consider that the relation of perceptual indiscriminability fundamentally is transitive, and that the non-transitive linear model of Figure 1 in fact results from the transitive product model of Figure 2 (by an operation inverse to the operation of layering). On that view, the "true" relation of indistinguishability behind the operator $\Box$ is the relation R' of the layered model, not the indistinguishability relation R of the linear margin model. But conversely, one could interpret the situation the other way around, and consider that the non-transitive relation R of the model of Figure 1 is the primitive relation.

The choice between these two options depends in part on what one takes to be the best representation of the notion of epistemic context, and then on what one takes "perceptual" to mean when one talks of "perceptual indistinguishability" between contexts. The layered model of Figure 2 gives a way of having an infinite set of distinct accessibility relations, one for each individual context, which is equivalent in turn to defining a single accessibility relation between richer contexts defined as ordered pairs, as in Halpern's semantics. In comparison, the margin model of Figure 1 gives a simpler representation of the notion of context: on our view, this representation gives a more intuitive rendering of the perceptual *experience* that subjects might have of a soritical series (like a series of pens ordered by height, or a series of hues cleverly shaded, and so on). The difference of granularity between the two notions of context reflects itself in a difference between two acceptations of the notion of "perception". When Halpern opposes "perception about sweetness" and "reports about sweetness", for instance, Halpern understands "perceptual" in a narrow sense, namely by reference to the subpersonal processes by which an agent organizes his basic sensory information. The example given by Halpern is that of a robot with sensors, whose perception of sweetness goes by unambiguous thresholds. By contrast, what Halpern describes as "reports about sweetness" is supposed to correspond to the qualitative experience the robot has of sweetness, which Halpern sees as potentially biassed and ambiguous, leaving room for non-transitivity in discrimination. This notion of "report" is in fact the qualitative notion of "perception" that we intend when we talk of perceptual indistinguishability.

---

same as "actually knowing", relative to a standard two-dimensional semantics in the style of Kamp's two-dimensional semantics for temporal logic in [7].

There should be no misunderstanding, consequently, when we assert that in our opinion, "perceptual indistinguishability" can be conceived as non-transitive: by this, we have in mind exactly the kind of indistinguishability which Halpern would associate to reports about perception. But the difference is that we can take such a non-transitive relation as primitive, without having to match it to a complex operator in order to describe knowledge. We do not think, therefore, that any epistemic uncertainty relation should necessarily be transitive. Concerning introspection, however, what Proposition 1 shows is that Halpern's approach in terms of product models and our approach in terms of a non-standard semantics follow essentially the same inspiration, by making higher-order knowledge depend only on the states that are in the immediate ken of the agent, and by avoiding spurious dependencies to alternatives that lie beyond those on which knowledge of the first-order supervenes.

4 Improving on the margins

So far, our approach to vagueness and introspection has been focused on modeling issues: given that the relation of perceptual indiscriminability does not seem to be transitive, how could we possibly recover introspection principles? Two strategies have been discussed: either the modal semantics has to be modified (our centered semantics), or the apparent intransitivity of the accessibility relation has to be factored out (Halpern's bimodal modeling). Now, Williamson does not only repudiate introspection principles because they do not hold in his margin for error semantics. He also offers a general argument against introspection that is meant to be theory-free and in particular independent of Kripke-style modeling of inexact knowledge.[13] What good is Centered Semantics – or Halpern's semantics for that matter – if there is a semantically neutral argument against introspection?

Williamson's argument rests on a scenario which is a quantitative variation on the pen and box situation we introduced at the beginning of this paper. The story goes like this. Mr Magoo sees a tree in the distance. He is pretty sure that that the tree is strictly less than k inches tall. However, Mr Magoo's knowledge obeys a margin for error principle: if the size of the tree had been between $k - \alpha$ and k inches, he would not have been able to tell that its size was less than k inches. We shall assume moreover that Mr Magoo has a reflective access to his knowledge: when he knows something, he knows that he knows it. Here is how Mr Magoo might improve on his

---

[13]We are indebted to an anonymous reviewer of our previous paper [1] for drawing our attention to this point, which was not addressed in [1].

woodsy knowledge (we shall use the notation $[s(t) < k]$ for "the size of the tree is less than k inches"; we here use K instead of $\square$ for the knowledge operator, in order to emphasize the fact that the reasoning is syntactic):

(i) $K[s(t) < k]$, by hypothesis

(ii) $K([s(t) \geq k - \eta] \rightarrow \neg K[s(t) \leq k])$, by reflection on the conditions of one's knowledge.

(iii) $KK[s(t) \leq k]$, by introspection and 1.

(iv) $K[s(t) < k - \eta]$ by 2, 3 and closure of knowledge under logical consequence.

By iterating this reasoning, we can show that Mr Magoo knows that the size of the tree is less than k' for arbitrary k'. This is clearly a *reductio*, and therefore one has to reject:

- either (ii) and the idea that Mr Magoo can know, by reflecting on the limitations of his visual abilities, that his margin for error is at least η for an arbitrarily small but fixed η,

- or (iii) and the idea that when Mr Magoo knows something, he knows that he knows it,

- or (iv) and the idea that Mr Magoo knows the logical consequence of what he knows.

Since truth is preserved under logical consequence, it seems that Mr Magoo can safely be assumed to be a good logician. According to Williamson, we should moreover grant Mr Magoo the ability to reflect on his visual limitations. So Williamson takes his argument as a *reductio* of (iii): introspection is not compatible with situations of inexact knowledge.

According to centered semantics, and similarly according to Halpern's bimodal approach (see [5], section 4.5),[14] (ii) is not valid. Note that since the logics associated with these semantics are normal, they satisfy (iv). Moreover, since they validate introspection, they satisfy (iii). The rejection of (ii) comes therefore as no surprise. In particular, in a fixed margin model with margin α, the principle of margin for error holds, but it does not hold that it is known. Let us consider the natural margin model $\mathcal{M}$ representing Mr Magoo's visual experience of trees. The model is $\mathcal{M} = \langle W, d, \alpha, V \rangle$. W is the

---

[14]In the following, the discussion will focus on CS rather than Halpern's setting. Because of the translations given in the previous section, this is no loss of generality.

set of possible sizes of the tree, say $W = \mathbb{R}^+$ if the tree could be r inches tall for any positive real number r. The distance between two worlds r and r' is $|r - r'|$. The margin α is a given number representing Mr Magoo's inability to distinguish between two trees whose sizes differ by less than α inches. Finally, V makes $s(t) < k$ true at r iff $r < k$. $[s(t) \geq k - \alpha] \rightarrow \neg K[s(t) \leq k]$ is CS-valid on $\mathcal{M}$ (it is true no matter what the size of the tree is). However, for given k and $\eta \leq \alpha$, $K([s(t) \geq k - \eta] \rightarrow \neg K[s(t) \leq k])$ does not hold in general, even with η much smaller than α.[15] For example, assuming that Mr Magoo's margin is 50 inches, $599 \nvDash_{CS} K([s(t) \geq 650 - 2] \rightarrow \neg K[s(t) \leq 650])$, because $(599, 648) \vDash_{CS} [s(t) \geq 650 - 2]$ and $(599, 648) \vDash_{CS} K[s(t) \leq 650])$. Thus, if the tree is 599 inches tall, CS predicts that Mr Magoo does not know that, if the tree is more than 648 inches tall, he does not know that it is less than 650 inches tall.

So, does Centered Semantics really makes inexact knowledge compatible with introspection? Maybe the price to pay is too high, because CS seems to make a very counterintuitive prediction by denying to Mr Magoo any kind of knowledge of the limitations of his visual abilities. Similarly, Williamson could claim that our account misses the point when it comes to addressing his argument against introspection, because it escapes the *reductio* by denying the much plausible assumption (ii). We shall now argue that Williamson's argument should be considered as a *reductio* of (ii) rather than of (iii), and that it does make sense to model a notion of knowledge for which (ii) fails.

As Williamson himself remarks, (ii) is all the more plausible than η can be taken to be arbitrarily small for the *reductio* to work. In particular η does not have to be α itself, and actually it does not seem epistemologically right to grant Mr Magoo the ability to know exactly his margin by mere reflection. Thus, (ii) does not amount to knowledge of one's margin for error: it just amounts to an ability to assign a lower bound to one's margin. Mr Magoo might not have any *a priori* access to his own margin for error. Still it seems that he can know by mere reflection that it is at least η, when η is reasonably small. And this seems to be enough to make (ii) a reasonable assumption. Or is it?

Let us make the issue more vivid. Suppose Mr Magoo engages in a discussion with his friend, the forest warden. The tree is 599 inches tall, Mr Magoo is not a professional, but he is not that bad at judging heights, say his margin for error is 50 inches. Mr Magoo does know that the tree is less than 650 inches tall, and he says so to the warden. Now the warden can ask him whether he cannot do better than that: "come on, Mr Magoo, if the

---

[15]This is a special feature of CS. On standard Kripke semantics, if a formula ϕ is true everywhere in a model, $K\phi$ is also true everywhere in that model.

tree was just 6 inches short of 650 inches, you could not tell me that it is no more than 650 inches tall!" The warden is right, and Mr Magoo, following the suggestion of his friend, engages in a reflection on the limitation of his eyesight. Eventually, he comes to know that the warden is right. He infers, rightly again, that the tree is less than 644 inches tall. But now his margin has changed: it is no more 50, but only 44. What happens if Mr Magoo's friend is actually a Williamsonian warden and makes the same point again and again? First, let us note that at some stage, the warden point will simply not be right, even granting than at all the previous stages Mr Magoo is entitled to improve on his knowledge by mere reflection. After 9 socratic rejoinders by the warden, Mr Magoo's margin will be no more than $50 - (9 \times 6) = 5$, and it will no more be true that Mr Magoo's margin for error is at least 6. Hence Mr Magoo will *not* be entitled to improve on his woodsy knowledge by considering that his margin is at least 6. Moreover, from an epistemic perspective, it is clear that Mr Magoo's ability to provide by truth-producing reflection a lower bound for his margin for error will decrease as the warden keeps insisting. Maybe the second time the warden asks, Mr Magoo can still be said to be entitled to consider that his margin is at least 3, given that his actual margin is now 44. On this scenario, the third time this process is iterated, Mr Magoo's margin is down to 41, and he might still be said to be entitled to consider that his margin is at least 1. But eventually, as the number n of iterations grows, Mr Magoo's reflective gain becomes smaller and smaller.

Let us go back to the abstract reasoning proposed by Williamson. A crucial point for Williamson's argument to go through is that it has to be possible to find a *fixed* value for η which can be reused in reiterating the inference. But, if the previous little story has it right, this is simply not possible! Let us assume that Mr Magoo has performed Williamson's inference once. Now he knows that $K[s(t) < k - \eta]$. Performing the reasoning once amounts to reducing Mr Magoo's margin, which is now $R(\alpha) = \alpha - \eta$, where the notation $R(\alpha)$ is meant to make explicit the functional dependency of Mr Magoos's reflective improvement on his visual limitations, represented by the value α. Now the sequence $\alpha, R(\alpha), R(R(\alpha)), ..., R^x(\alpha), ...$ has to have a positive – presumably non-zero – lower bound: margins cannot become negative, just as the warden cannot make his point rightly 10 times in a row, because on the ninth time, Mr Magoo's actual margin is less than 6. By simple mathematics, this implies that there is no *fixed* value η, no matter how small, such that we can take $R^{x+1}(\alpha)$ to be $R^x(\alpha) - \eta$. This is why we think that the seemingly innocent assumption (ii), and the idea that Mr Magoo can know, by reflecting on the limitations of his visual abilities, that his margin for error is at least η for an arbitrarily small but fixed η, should

be rejected.[16]

Note that the existence of a lower bound does not contradict the weaker assumption that Mr Magoo's can perform Williamson's reasoning as many times as he wants, *and* that by doing so he indefinitely improves on his knowledge. More precisely, we need not settle the question whether the limit of the sequence is actually reached or not, *i.e.* whether there is an x such that $R^x(\alpha) = R^{x+1}(\alpha)$. In other words, we can grant to Williamson that, at each step, it is possible to find an η such that (ii) is correct, though we do not grant the much stronger claim that there is an η such that, at each step, (ii) is correct. Actually, we could see how Williamson's justification for (ii) can be construed as supporting the weaker claim. But the unwelcome consequence of Williamson's original arguments seems to us to provide a good reason to reject the stronger claim.

By definition, the sequence $\alpha, R(\alpha), R(R(\alpha)), ..., R^x(\alpha), ...$ is monotone decreasing, that is $R^{x+1}(\alpha) \leq R^x(\alpha)$. Moreover it is bounded by zero. By simple mathematics again, it has a limit. In other words, there is an A such that

$$\lim_{x \to +\infty} R^x(\alpha) = A$$

Intuitively, A corresponds to the limitation of Mr Magoo's ability to distinguish heights, given both his visual limitations *and* his ability to engage himself repeatedly in the previous reflections on the limitations of his knowledge. Accordingly, $R(A)$ should be taken to be A itself: it corresponds to the maximal amount of information that Mr Magoo is capable to extract from his visual experience of seeing a tree in the distance. In that case, assumption (ii) is no more an option: by construction, we have reached the point where Mr Magoo's ability to improve on his prior knowledge has come to an end.

These considerations vindicate our initial claims that Williamson's argument should be considered as a *reductio* of (ii), and that it does make sense to model a notion of knowledge for which (ii) fails. (ii) is faulty because the assumption that η can be set to a fixed value is not correct. Moreover, we can grant Williamson's point that, starting with a given amount of purely perceptual knowledge, Mr Magoo can gain new knowledge by reflection. But this process reaches a limit – as we have just seen, pretty much in the mathematical sense of the word. Knowledge at the limit is both perceptual and inferential: it comes from what is perceived and from (idealized) reflection on

---

[16]As the reasoning is repeated, the justifications for (ii) cannot remain the same, because Mr Magoo's knowledge is less and less based on perception and more and more inferential, as his margin shrinks. As a consequence, there is no reason in the first place to assume that η can be given a uniform value in (ii), since the justifications for the various instances of (ii) are not themselves uniform.

perceptual limitations. And for this notion of knowledge at least, the failure of all instances of (ii) even for very small values of η is perfectly reasonable, precisely because, by hypothesis, we consider what the agent can be said to know on account that he has maximally exploited her ability to reflect on her perceptual limitations.

5 Conclusion

In this paper we have compared three different models of inexact knowledge in epistemic logic. The first is Williamson's model, which rests on non-transitive and non-euclidian relations but invalidates positive and negative introspection. A second model is Halpern's explicit two-dimensional logic, in which non-transitive (and non-euclidian) relations of epistemic uncertainty are conceived as the product of more basic transitive relations. Our approach, finally, while close in spirit to Halpern's approach, can be seen to provide a compromise between his approach and that of Williamson, since by modifying the semantics, it allows us to preserve introspection directly over non-transitive models. Fundamentally, as we have seen, Halpern's account and ours agree on the idea that knowledge can be inexact and yet support standard introspection principles. From a conceptual point of view, however, Centered semantics is a way of implicitly contextualizing relations of epistemic indiscriminability characteristic of vagueness. Moreover, while Centered Semantics arguably gives a model of first-order vagueness without higher-order vagueness, we have seen that it belongs to a natural hierarchy of semantics for finite higher-order vagueness.

In the last part of this paper, we have seen that our semantics, like Halpern's, does not allow the agent to know her margin for error systematically, on pain of contradiction. Williamson's argument, as we construe it, shows that it is not an accidental feature of Centered Semantics or of Halpern's bimodal system. However, this does not mean that we do not grant the agent the ability to reflect on the limitations of her perceptual apparatus and to approximate her margins for error. Rather, our point is that this process is to be conceived as a dynamic process of more and more refined approximations, which will eventually reach a limit, precisely because each step of reflection on one's margin is likely to be a step that withdraws from the initial margin. We have not yet provided a formal rendering of the process by which an agent gains new knowledge by reflecting on her margins. Dynamic logic would be the natural tool for doing so, however, since starting from purely perceptual data, the agent would update her prior knowledge by approximations of her margin for error, until she reaches a fixed point corresponding to a stable state of knowledge.

References

[1] BONNAY D. & P. ÉGRÉ (2006), A non-standard Semantics for Inexact Knowledge with Introspection, forthcoming in S. Artemov & R. Parikh (eds), *Proceedings of the ESSLLI 2006 Workshop on Rationality and Knowledge*, Malaga.

[2] BONNAY D. & P. ÉGRÉ (forthcoming), Margins for Error in Context, in M. Garcia-Carpintero & M. Kölbel (eds.), *Relative Truth*, Oxford University Press.

[3] FARA GRAFF D. (2001), Phenomenal Continua and the Sorites, *Mind*, vol. 110(440) pp. 905-935. Originally published under the name "Delia Graff".

[4] FARA GRAFF D. (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".

[5] HALPERN J. (2004), Intransitivity and Vagueness, *Proceedings of the Ninth International Conference on Principles of Knowledge Representation and Reasoning (KR 2004)*, pp. 121-129.

[6] HORGAN T. (1994), Robust Vagueness and the Forced-March Sorites Paradox, *Philosophical Perspectives*, 8: *Logic and Language*, 159-88.

[7] KAMP H. (1971), Formal Properties of "Now", *Theoria* 37, 227-273.

[8] RABINOWICZ W. & K. SEGERBERG (1994), Actual Truth, Possible Knowledge, *Topoi* 13, 101-15.

[9] RAFFMAN D. (2000), Is Perceptual Indiscriminability Nontransitive?, *Philosophical Topics*, 28, 1, *Vagueness*.

[10] SHAPIRO S. (2003), Vagueness and Conversation, in J.C. Beall (ed.), *Liars and Heaps*, pp. 39-72, Oxford UP.

[11] WILLIAMSON T. (1992), Inexact Knowledge, *Mind*, 101, pp. 217-42.

[12] WILLIAMSON T. (1994), *Vagueness*, Routledge.

[13] WILLIAMSON T. (1999), On the Structure of Higher-Order Vagueness, *Mind*, vol. 108, 127-43.

[14] WILLIAMSON T. (2000), *Knowledge and its Limits*, Oxford University Press.

Appendix: Token Semantics

Token Semantics is a generalization of Centered Semantics by which only worlds that are n steps away from the actual world can be visited for a given n. In Token Semantics, the evaluation of a formula is parameterized by a number n of tokens: each move along the accessibility relation costs one token, and once all tokens have been spent, no worlds that are essentially further away can be accessed.

Formally, satisfaction is defined with respect to a sequence of worlds and a number of tokens: q is short for an arbitrary sequence of worlds, qw for an arbitrary sequence augmented with w, and n is an arbitrary number of tokens.

Definition 3. *Token satisfaction:*

(i) $\mathcal{M}, qw \vDash_{\text{TS}} p \ [n]$ iff $w \in V(p)$.
(ii) $\mathcal{M}, qw \vDash_{\text{TS}} \neg\phi \ [n]$ iff $\mathcal{M}, qw \nvDash_{\text{TS}} \phi \ [n]$.
(iii) $\mathcal{M}, qw \vDash_{\text{TS}} (\phi \wedge \psi) \ [n]$ iff $\mathcal{M}, qw \vDash_{\text{TS}} \phi \ [n]$ and $\mathcal{M}, qw \vDash_{\text{TS}} \psi \ [n]$.
(iv) $\mathcal{M}, qw \vDash_{\text{TS}} \Box\psi \ [n]$ iff
       • $n \neq 0$ and for all w' such that wRw', $qww' \vDash_{\text{TS}} \psi \ [n-1]$
       • Or $n = 0$ and $q \vDash_{\text{TS}} \Box\psi \ [1]$.

Definition 4. *Let n be such that $1 \leq n \leq \omega$ (we assume $\omega - 1 = \omega$). TS(n) is the modal semantics defined by the following satisfaction relation: $\mathcal{M}, w \vDash_{\text{TS}(n)} \phi$ iff $\mathcal{M}, w \vDash_{\text{TS}} \phi \ [n]$.*

By definition, a formula is TS(n)-valid if it is TS(n)-true at every world of every model. CS is equivalent to TS(1), and standard Kriple semantics is equivalent to TS(ω).

We have seen that CS validates (4) and (5). Similarly, TS(n) validates the following schemata:

$(4n)$ $\square^n p \rightarrow \square^n \square p$

$(5n)$ $\Diamond^n p \rightarrow \Diamond^{n-1} \square \Diamond p$

where $\square^n$ is short for $\square...\square$, n times. In [1] we provide a complete axiomatization for the semantics TS(n), which rests on a strengthened version of the axioms $(4n)$ and $(5n)$. With respect to standard Kripke semantics, these strengthened schemas axiomatize the class of n-transitive and n-euclidian frames, where an n-transitive frame is a frame satisfying:

$\forall x_1, ..., x_{n+2}\ ((x_1 R x_2 \wedge ... \wedge x_{n+1} R x_{n+2}) \rightarrow x_n R x_{n+2})$

and an n-euclidian frame is a frame satisfying:

$\forall x_1, ..., x_{n+2}\ ((x_1 R x_2 \wedge ... \wedge x_n R x_{n+1} \wedge x_n R x_{n+2}) \rightarrow x_{n+2} R x_{n+1})$

DENIS BONNAY
Département d'Etudes Cognitives
Ecole Normale Supérieure, 29, rue d'Ulm
75005 Paris - France
denis.bonnay@univ-paris1.fr

PAUL EGRÉ
Institut Jean-Nicod
CNRS
ENS-DEC - 29, rue d'Ulm
75005 Paris - France
paulegre@magic.fr

21