CANADIAN JOURNAL OF PHILOSOPHY, 2016
VOL. 46, NOS. 4–5, 803–835
http://dx.doi.org/10.1080/00455091.2016.1201878

Routledge
Taylor & Francis Group

Epistemicism and modality

Juhani Yli-Vakkuri

CSMN, University of Oslo, Oslo, Norway

ABSTRACT

What kind of semantics should someone who accepts the epistemicist theory of vagueness defended in Timothy Williamson's *Vagueness* (1994) give a definiteness operator? To impose some interesting constraints on acceptable answers to this question, I will assume that the object language also contains a metaphysical necessity operator and a metaphysical actuality operator. I will suggest that the answer is to be found by working within a three-dimensional model theory. I will provide sketches of two ways of extracting an epistemicist semantics from that model theory, one of which I will find to be more plausible than the other.

ARTICLE HISTORY Received 11 June 2016

KEYWORDS Vagueness; modality; epistemicism; Timothy Williamson

What kind of semantics should someone who accepts the epistemicist theory of vagueness defended in Timothy Williamson's *Vagueness* give a definiteness operator? To impose some interesting constraints on acceptable answers to this question, I will assume that the object language also contains a metaphysical necessity operator and a metaphysical actuality operator. The question, then, concerns the semantics of the interaction of vagueness, as expressed by 'definitely,' and metaphysical modality, as expressed by 'necessarily' and 'actually,' from an epistemicist point of view. I will argue that the kind of two-dimensional (2D) semantics standardly used for investigating the interaction of 'necessarily' and 'actually' cannot handle the three-way interaction between these and 'definitely,' given epistemicism (or indeed supervaluationism). I will suggest that the answer is to be found by working within the three-dimensional (3D) model theory I have developed with Jon Litland. I will provide sketches of two ways of extracting an epistemicist semantics from that model theory, one of which I will find to be more plausible than the other.

CONTACT Juhani Yli-Vakkuri t.j.yli-vakkuri@csmn.uio.no

1. According to the epistemicist theory of vagueness (*epistemicism*) defended by Williamson (1994), our ignorance of borderline matters — such as of how many grains of sand it takes to make a heap — has a special source, and that source of ignorance or, perhaps better put, that obstacle to knowledge, is all there is to vagueness. For it to be borderline, whether *S* is simply for there to be a special kind of obstacle to us (and to creatures relevantly like us) knowing whether *S*. Williamson's account of this obstacle to knowledge famously makes use of the thesis that, whenever it is borderline whether *S*, the sentence '*S*' is, to use John Hawthorne's term,[1] *semantically plastic* in that, if the global patterns of use involving some of the words occurring in '*S*' had been ever so slightly different in certain ways, then '*S*' would have had a semantic value different from its actual semantic value. (I will henceforth refer to this thesis as *Semantic Plasticity*.) The rough idea of the explanation is that we are insensitive in our judgments to the ways in which certain slight differences in global patterns of language use make a difference to the semantic values of sentences, and this insensitivity explains why we are not in a position to know whether *S* whenever it is borderline whether *S*. The semantic values that the borderline sentence '*S*' could easily have had had the global patterns of use of some of the words occurring in '*S*' been ever so slightly different in certain ways include both true and false semantic values. This, in turn, gives rise to close possibilities of error — possibilities that are 'close' in a sense that matters to knowledge, and that are therefore incompatible with knowing whether *S*.

Epistemicism appears to be the simplest and most conservative theory of vagueness on offer. Its main advertised virtues are that it enables us to keep classical logic and a disquotational conception of truth,[2] and that, '[a]t least in its simplest form, epistemicism can take over and reinterpret the formal apparatus of supervaluationism,'[3] thereby using the familiar formal apparatus of possible-worlds model theory to investigate the logic and semantics of the operators we use for theorizing about vagueness: in the first instance, the definiteness operator, in terms of which the notion of a borderline case can be defined. (It is borderline whether *S* iff it is neither definite that *S* nor definite that it is not the case that *S*.)[4]

From the beginning, epistemicism faced the incredulous stare. During the first decade or so since the publication of *Vagueness*, there seemed to be a broad consensus that the choice epistemicism presented us with was one between some kind of intuitive plausibility (which it lacked) and theoretical virtues (which it did not).[5] Recently, however, epistemicism's theoretical bona fides have been called into question. Since 2006, certain philosophers — in particular, John Hawthorne, Ofra Magidor, Stephen Kearns, and Michael Caie[6] — have begun to engage with the details of Williamson's proposal. In doing so they have discovered what appear to be quite serious problems with the proposed account of ignorance of borderline matters in terms of Semantic Plasticity, with epistemicist interpretations of supervaluationism's formal (model-theoretic) apparatus, and

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

with the compatibility of these with a commitment to disquotational truth. So far, these challenges have gone unanswered. This paper represents a start — but only a start — in attempting to answer them.

I should warn the reader immediately that no review of the recent literature will be provided, although some points of contact with the papers cited above will be noted. I will largely be concerned with challenges to the Williamsonian epistemicist package that are different from, although closely related to, those discussed in the recent critical literature. The main problem I will be dealing with is that of giving an epistemicist semantics for a language that contains a definiteness operator alongside a metaphysical necessity operator and a metaphysical actuality operator. The difficulties one faces in attempting to carry out this task parallel in various ways the difficulties Williamson's recent critics have noticed. This is not particularly surprising, because the critics, with the exception of Caie, set up their problems using the metalinguistic notion of a sentence (or a sentence in a context, or an utterance of a sentence) being borderline, as opposed to an object language borderline operator (which is definable in terms of a definiteness operator, as indicated above). These are related to each other in obvious ways: in particular, in that the sentence '*S*' is borderline iff it is borderline whether *S*. Different issues arise when the borderline (or definiteness) operator occurs within the scope of a modal operator, but many of the problems discussed in this paper do not essentially involve such embeddings.

2. Let us return to Semantic Plasticity. My description in Section 1 of Williamson's explanation of ignorance of borderline matters in terms of Semantic Plasticity was deliberately vague (in the ordinary sense). This is because much in the recent literature and in this paper turns on just what the correct account of ignorance of borderline matters in terms of Semantic Plasticity is, if indeed there is one. For example, I left it open whether the kind of sentential semantic value mentioned in Semantic Plasticity is a proposition or something else. But we can begin with the assumption that propositions are the relevant kind of semantic value, since they clearly play that role in Williamson's own expositions of his explanation of ignorance of borderline matters.[7] If there is to be an explanation in terms of Semantic Plasticity of why we cannot know whether *S* when it is borderline whether *S*, that explanation must posit close worlds in which '*S*' expresses a false proposition, as well as close worlds in which '*S*' expresses a true proposition. Only close possibilities of error are incompatible with knowledge; close non-actual possibilities in which one judges something that is merely actually false are not, and neither are actual possibilities in which one judges something that is merely counterfactually false. On the other hand, the relevant possibilities of error must match actuality in certain respects: the existence of a close world in which, e.g. 'Tim is thin' expresses a false proposition other than the proposition it actually expresses and in which Tim's precise physical measurements are different than they actually are may be incompatible with one's knowing that Tim is thin, but

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

it is not an example of the kind of obstacle to knowledge presented by vagueness. Much of the recent critical literature can be read as suggesting that these two demands on an explanation of ignorance of borderline matters in terms of Semantic Plasticity are difficult to reconcile with each other and with the other commitments of Williamson's epistemicism: in particular with disquotational truth and an epistemicist reinterpretation of supervaluationism's model-theoretic apparatus. In fact, Caie appears to think that the two commitments force a dilemma on Williamson's kind of epistemicism, both forks of which lead to clearly incorrect conclusions about what is borderline.

One way of throwing the difficulties I have just gestured at into particularly sharp relief is to ask what kind of a semantics someone committed Williamson's kind of epistemicism should give a definiteness operator. I will pursue this line of inquiry, and I will take Caie's recent discussion as my starting point, although his discussion, unlike mine, is not conducted in an explicitly model-theoretic framework. My discussion will also differ from Caie's (among other respects) in that I will ask not only what kind of semantics the epistemicist should give a definiteness operator — which I'll designate by '$\triangle$' — but a definiteness operator in combination with a metaphysical necessity ($\Box$) and actuality (A) operator. It will turn out that the inclusion of $\Box$ and A in the object language makes it possible to produce reasonably snappy counterexamples to certain semantic proposals that would otherwise require more elaborate counterexamples.

3. Let us begin by drawing a distinction between a *model theory* and a *semantics*.

A *semantics* for a language, as I will use the term in this paper, is an appropriately informative specification, in a metalanguage, of the function that assigns to each sentence of the language, in each context of utterance, the proposition the sentence expresses in that context. I will assume that propositions are (or are represented by — a distinction of which I will not be making heavy weather) functions from metaphysically possible worlds to truth values. In particular, I will assume that a proposition p is the function that assigns truth to a world w if p is true in w and otherwise assigns falsehood to w. (I make this assumption in order to simplify the discussion: nothing in this paper turns on whether propositions are more fine-grained than this.) When dealing with a suitably simple language — as in this paper — we can assume that contexts of utterance are also nothing more than metaphysically possible worlds. Metaphysically possible worlds (or simply 'worlds'), then, play two roles in a semantics: the role of contexts and the role of the entities with respect to which propositions are evaluated for truth. I will follow one standard nomenclature (that of Kaplan 1977) in calling worlds *circumstances* when they play the second role. A semantics for a language, then, is an appropriately informative specification of a function from the sentences of the language to functions from contexts to functions from circumstances to truth values. Equivalently, we can think of a semantics for a language as associating each of its sentences ϕ with a two-dimensional semantic value that is

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

its *total semantic profile*: a set *P* of ordered pairs of worlds such that $\langle w, v \rangle \in P$ iff the proposition ϕ expresses in *w* is true in *v*.

Three further remarks about the present conception of a semantics are in order.

First, we will have to think of at least some semantic features of the expressions of the object language as being fixed independently of context. At least we will have to assume that the conventional meanings — in Kaplanian terms, *characters* — of the logical constants, which in the present setting are $\triangle$, $\Box$, *A*, and the truth-functional connectives, do not vary with context.[8] Thus, for example, we assume that, for any context *w*, the proposition $\neg\phi$ expresses in *w* will be true in a circumstance iff the proposition ϕ expresses in *w* is false in that circumstance, regardless of how (and whether) the expression $\neg$ is used in *w*. I'll leave it an open question what else is fixed. In various other settings it would be natural to follow Kaplan in assuming that the characters of all object language expressions are fixed independently of context. If we did so, we should think of the total semantic profile of a sentence as simply being (or representing) its character. However, certain peculiar features of the task of giving a semantics for a language containing $\triangle$ might be thought to rule out this assumption, so I won't make it. I will return to this theme in Section 4.

Second, it will have to be decided what the proposition expressed by ϕ in *w* is when ϕ is not used in *w*. We will have to resolve to use the term 'expresses' in a technical sense according to which the proposition an atomic sentence ϕ expresses in *w*, in the technical sense, when ϕ is used in *w*, is the proposition ϕ expresses in *w* in the ordinary sense, but the proposition an atomic sentence ϕ expresses in *w*, in the technical sense, when ϕ is not used in *w*, is something else — for example, it might be the proposition that is false in every world. (For non-atomic sentences the proposition is determined by the recursion clauses of the semantics.)[9] It makes no difference how this question is decided, so I won't decide it.

(A related idealization: I assume that, when ϕ *is* used in *w*, ϕ expresses a unique proposition in *w*. Thus, I idealize away, *inter alia*, ambiguity, semantically defective uses of sentences, and dimensions of indexicality other than sensitivity to the world of the context.)

Third, a semantics may, and in almost any interesting case will, do more than give an appropriately informative description of the total semantic profile of each sentence of the object language. It is a natural and widely endorsed methodological precept (which is, however, rarely observed in practice[10]) that a semantics should specify an assignment of semantic values to sentences in a way that is, in the standard sense of the word, *compositional*, and that these semantic values either are or determine the sentences' total semantic profiles. Often the compositional semantic values can only determine, rather than be, the total semantic profiles. And this is how things go, as we will see, with a language containing $\triangle$, $\Box$, and *A*, when $\triangle$ is given an epistemicist interpretation:

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

the compositional semantic values of its sentences must be three-dimensional rather than two-dimensional. (This is also true of supervaluationist-friendly semantics for $\triangle$, $\Box$, and *A*: see Litland and Yli-Vakkuri, 2016.) But in such cases we can assume that each context supplies a unique value for each parameter, and we can still take the total semantic profile of a sentence to be a set of ordered pairs of contexts and circumstances, since the contexts in such pairs will determine values for all of the parameters that it takes to assign a sentence a unique truth value at each circumstance.

A *model theory*, on the other hand, is a definition of a relation designated as *truth in a model*, or, in the cases that will interest us, *truth at a point of evaluation in a model*, supplemented by two further definitions: a *model* is defined as a certain kind of set, and a relation $\models$ of *logical consequence* is defined as (e.g.) truth-preservation at every (proper) point of evaluation of every model. These definitions supplement some background set theory — ZFC, say — so that anything proved by means of them will be a theorem of that set theory. Among the interesting claims that might be proved or disproved in this way is that the set of sentences designated as *valid* (i.e. the logical consequences of the empty set) is recursively axiomatizable ('completeness'). Such results may be interesting on their own, even if no bridge principles are supplied to connect the model theory to a semantics in the above sense — or to semantics in any natural sense.

The development of a model theory, however, is typically guided by a semantic picture, in a broader sense of 'semantic.' The broadest contours of such a picture are often the following. A model is (or represents) a way of interpreting, in some sense, all of the non-logical expressions of the object language. Furthermore, when models provide points of evaluation, but do not provide a privileged 'actual' point,[11] the 'proper' points of evaluation are naturally taken to be (or represent) contexts. Logical consequence, then, is truth-preservation in all contexts under all ways of interpreting the object language's non-logical expressions.

If we assume this broad-outlines picture, as I will do, we can go on to ask questions about the model that is (or represents) the actual or correct interpretation of the non-logical expressions of the object language — the *intended model*, as it is usually called. If the models come with representatives of contexts as well as circumstances, then a suitable specification of the intended model will be a semantics.[12]

The standard approach to the model theory of languages with $\triangle$ is that associated with supervaluationism.[13] In the simplest case, where we are dealing with a propositional language whose only non-truth-functional logical operator is $\triangle$, a supervaluationist model is a creature familiar from texts on modal logic: a triple $\mathfrak{U} = \langle W, R, [\![.]\!] \rangle$, where *W* is a non-empty set (of 'valuations'), *R* a reflexive relation on *W* (perhaps satisfying some further conditions), and $[\![.]\!]$ a function from the atomic sentences to subsets of *W*. An atomic sentence ϕ is defined as true in $\mathfrak{U}$ at $w \in W$ iff $w \in [\![\phi]\!]$; we have the usual truth definition clauses for the

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

truth-functional connectives; and $\triangle\phi$ is defined as true in $\mathfrak{U}$ at $w \in W$ iff, for all v such that wRv, ϕ is true in $\mathfrak{U}$ at v. Logical consequence can be defined in several ways, which will not be of interest here. The question I want to ask, rather, is this: how, if at all, could we extract an *epistemicist* semantics for a language with $\triangle$ from the supervaluationist's model theory?

Some remarks by Williamson suggest optimism on this score:

> At least in its simplest form, epistemicism can take over and reinterpret the formal apparatus of supervaluationism, with a Tarskian conception of truth. To say that a valuation *V** is admissible by [*R*-related to] a valuation *V* is now to say that *V** is indiscriminable from *V* in the sense indicated. (Williamson 2003, 710)

On one natural way of extrapolating an epistemicist semantics from these remarks, $\triangle$, as used in a given world w, is, in effect, a device for generalizing over assignments of propositions-*cum*-semantic-values to sentences that obtain in worlds that are *close* to w in such a way that global patterns of the use of language that obtain in them differ from those obtaining in w at most in such minor ways that we are insensitive to any differences these differences in use may make to semantic value — insensitive, furthermore, in a way that makes the semantic value assignment that obtains in any world close to w indiscriminable (to creatures like us) from the semantic value assignment that obtains in w. Caie (2012) uses the term 'semantic indiscriminability' for this kind of closeness, and I will adopt his term: I will assume, for now, that two worlds are close to each other in the sense relevant to an epistemicist interpretation of $\triangle$ iff they are semantically indiscriminable from each other in the rough-and-ready sense just indicated.

(Note that the supervaluationist's models, as described above, do not contain any representatives of propositions: they simply assign truth values to atomic sentences relative to valuations. It is, however, natural to think of the valuations as representing ways of assigning propositions to atomic sentences, and to think of the truth value of an atomic sentence ϕ relative to a valuation v as the truth value determined by the proposition ϕ expresses relative to v.)

Caie, however, is not optimistic that anything like the above sketch can be developed into a satisfactory epistemicist semantics for $\triangle$. In fact, Caie goes as far as to suggest that the difficulties he encounters in attempting to extrapolate an epistemicist semantics for $\triangle$ from the quoted remarks by Williamson support the conclusion that 'one cannot provide an adequate account of what it is for a case to be borderline by appealing to facts about our inability to discriminate our actual situation from nearby counterfactual situations in which our language use differs in subtle ways.'[14]

Let us see, then, why Caie thinks this.

Caie observes that there appear to be two inequivalent accounts of what it takes for $\triangle\phi$ to be true in a world (considered as a context, I assume) that are consistent with Williamson's account of ignorance of borderline matters. According to the account Caie calls 'otherworldly,' $\triangle\phi$ is true in a world w (considered as a context) iff, for every world v close to w, the proposition-*cum*-semantic value

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

ϕ has or expresses in *v* (considered as a context) is true in *v* (considered as a circumstance). According to the other account, which Caie calls 'actualistic,' $\triangle\phi$ is true in a world *w* (considered as a context) iff, for every world *v* close to *w*, the proposition-*cum*-semantic value ϕ has or expresses in *v* (considered as a context) is true in *w* (considered as a circumstance). Caie claims that both answers 'face crippling defects.'[15]

One thing to note about Caie's otherworldly and actualistic accounts is that neither provides enough information for a semantics for a language with $\triangle$, even against a background theory that specifies the total semantic profile of every sentence containing no occurrences of $\triangle$. Each account only purports to say what it takes for $\triangle\phi$ to be true in a single world *w* — considered as a context, I have assumed — which is to say, what it takes for the proposition expressed by $\triangle\phi$ in a given world to be true *in that world*. This is not enough to specify the proposition expressed by $\triangle\phi$ in any context. Now, of course, if Caie is right that absurd consequences follow from the what the actualistic and otherworldly accounts say about what it takes for the proposition expressed by $\triangle\phi$ in a world to be true in that same world, then there is no need to consider separately the semantic proposals consistent with them which specify what it takes for the proposition expressed by $\triangle\phi$ in one world to be true in an arbitrary world. However, I do not find Caie's arguments against the otherworldly and actualistic accounts entirely decisive,[16] and I think we can learn something by regimenting Caie's two options using standard model-theoretic resources. In fact, as promised, I will consider not only what an epistemicist would have to say about the semantics of $\triangle$, but also of its interaction with $\Box$ and *A*, if the semantic options are limited to those consistent with the otherworldly and actualistic options.

4. Let L_{VM} be a propositional language with an infinite stock of atomic sentences, the usual truth-functional connectives, with $\neg$ and $\wedge$ taken as primitive, $\Box$ ('necessarily'), *A* ('actually'), and $\triangle$ ('definitely'), and no other logical constants, with the usual formation rules and metalinguistic abbreviations. Thus, in particular, '$\Diamond\phi$' abbreviates '$\neg\Box\neg\phi$' and '$\nabla\phi$' abbreviates '$\neg\triangle\phi \wedge \neg\triangle\neg\phi$.' Let L_M be the language that results from dropping $\triangle$ from L_{VM}. L_{VM} is the language whose semantics this paper will, for the most part, be concerned with. Languages with greater expressive resources will also be briefly considered.

The standard model-theoretic treatment of L_M is two-dimensional: it interprets sentences as sets of two-dimensional points of evaluation drawn from the Cartesian product of a non-empty set *W* and itself. In what follows I will depart from that standard treatment by including in each model an 'accessibility' relation *R* which is *not* used for interpreting $\Box$; rather, *R* will later be used, just as in supervaluationist model theory, for interpreting $\triangle$ in L_{VM}, so *R* will be idle as far as the treatment of L_M is concerned.

So let us say that a *2D model* is a triple $\mathfrak{U} = \langle W, R, [\![.]\!] \rangle$, where *W* is a non-empty set, *R* a reflexive and symmetric relation on *W*, and $[\![.]\!]$ a function from the

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

atomic sentences of L_M (which are also the atomic sentences of L_{VM}) to subsets of $W \times W$. The truth value $|\phi|^{\mathfrak{U}}_{w,v} \in \{0, 1\}$ of an L_M-sentence ϕ at a point of evaluation $\langle w, v\rangle \in W \times W$ in $\mathfrak{U}$ is defined as follows, where α is any atomic sentence and ϕ and ψ are any sentences.

(i) $|\alpha|^{\mathfrak{U}}_{w,v} = 1$ iff $\langle w, v\rangle \in [\![\alpha]\!]$

(ii) $|\neg\phi|^{\mathfrak{U}}_{w,v} = 1$ iff $|\phi|^{\mathfrak{U}}_{w,v} = 0$

(iii) $|\phi \wedge \psi|^{\mathfrak{U}}_{w,v} = 1$ iff both $|\phi|^{\mathfrak{U}}_{w,v} = 1$ and $|\psi|^{\mathfrak{U}}_{w,v} = 1$

(iv) $|\Box\phi|^{\mathfrak{U}}_{w,v} = 1$ iff for all $u \in W$, $|\phi|^{\mathfrak{U}}_{w,u} = 1$

(v) $|A\,\phi|^{\mathfrak{U}}_{w,v} = 1$ iff $|\phi|^{\mathfrak{U}}_{w,w} = 1$

Let's say that a point of evaluation is *proper* iff it is of the form $\langle w, w\rangle$, and that ϕ is *true in w in* $\mathfrak{U}$ ($|\phi|^{\mathfrak{U}}_{w} = 1$) iff ϕ is true at the proper point of w, i.e. at $\langle w, w\rangle$, in $\mathfrak{U}$. Logical consequence is defined as truth-preservation at every proper point of evaluation of every model: that is to say, for any set of sentences Γ and any sentence ϕ, $\Gamma \vDash \phi$ iff, for all models $\mathfrak{U} = \langle W, R, [\![.]\!]\rangle$ and all $w \in W$, $|\phi|^{\mathfrak{U}}_{w} = 1$ if, for all $\gamma \in \Gamma$, $|\gamma|^{\mathfrak{U}}_{w} = 1$. And a sentence ϕ is said to be valid ($\vDash \phi$), as usual, iff it is a consequence of the empty set.

(Below I will tend to drop the model superscript from the expressions '$|\phi|^{\mathfrak{U}}_{w,v}$' and '$|\phi|^{\mathfrak{U}}_{w}$,' writing simply '$|\phi|_{w,v}$' or '$|\phi|_{w}$,' when it is clear from the context what generalization over models is meant or which model is being referred to.)

The above definitions give us the logic of 'real-world validity.'[17] The distinctive feature of this logic is that the principle of Necessitation (if $\vDash \phi$ then $\vDash \Box\phi$) fails. For example, $\phi \equiv A\phi$ is true at every proper point of evaluation of every model, so is valid, but whenever $\neg\Box\phi \wedge \neg\Box\neg\phi$ is true at a proper point in a model, $\Box(\phi \equiv A\phi)$ is false at that point in that model, so is not valid.

The extraction of a semantics for L_M from the 2D model theory, according to the standard view, which we largely owe to Kaplan (1977), proceeds in a straightforward way. The idea is that, in the intended model, W is the set of all metaphysically possible worlds, and the worlds in a point of evaluation $\langle w, v\rangle$ play the two roles described in Section 3: w is a context, and v is a circumstance. $|\phi|_{w,v} = 1$ iff ϕ is true at $\langle w, v\rangle$ in the sense of Section 3: i.e. iff the proposition ϕ expresses in the context w is true in the circumstance v.[18] Thus, the proposition expressed by ϕ in the context w is the function $\lambda v|\phi|_{w,v}$. Since ϕ is true *simpliciter* iff the proposition ϕ actually expresses is actually true, ϕ is true *simpliciter* iff $|\phi|_{@} = 1$, i.e. iff $|\phi|_{@,@} = 1$, where '@' is used, as usual, as a name for the actual world.

It's worth noting — to return to a theme from Section 3 — that the particular version of this semantics that is favored by Kaplan will not do for epistemicist (or, indeed, supervaluationist) applications. According to the orthodox Kaplanian version, the function $\lambda w\lambda v|\phi|_{w,v}$ is the character of ϕ. This would have the consequence that no sentence with a constant character that determines, in every

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

context, the same non-contingent proposition is borderline. For suppose that ϕ is such a sentence. Then ϕ has the same truth value at every point of evaluation, and consequently $\triangle\phi \vee \triangle\neg\phi$ is true at every point. But this is clearly wrong. A sentence with a constant character that expresses, in every context, the same non-contingent proposition could easily be borderline. For example, there are vague identity sentences with proper names flanking the identity predicate. Suppose, for example, that it is borderline whether *S* and that we introduce the name '*N*' with the stipulation that '*N*' refers to Kit Fine if *S* and to Tim Williamson otherwise;[19] then it will be borderline whether *N* = Kit, and it will be borderline whether *N* = Tim. However, given the standard assumptions that names are rigid designators in the sense of Kripke (1980) and are not indexicals, the sentences '*N* = Kit' and '*N* = Tim' will have constant characters each of which determines, in each context, the same non-contingent proposition.[20]

Now let us turn to L_{VM}. To extend the above model theory to deal with L_{VM}, we will have to add a clause for $\triangle$ to the definition of truth at a point in a model. If we restrict our options to clauses that are compatible with one or another of Caie's otherworldly and actualistic accounts of the truth of $\triangle\phi$ at a proper point, we have four options. The otherworldly account is consistent with each of the following.

(O_1) $|\triangle\phi|^{\mathfrak{U}}_{w,v} = 1$ iff, for all u such that vRu, $|\phi|^{\mathfrak{U}}_{u,u} = 1$.

(O_2) $|\triangle\phi|^{\mathfrak{U}}_{w,v} = 1$ iff, for all u such that wRu, $|\phi|^{\mathfrak{U}}_{u,u} = 1$.

And the actualistic account is consistent with each of the following.

(A_1) $|\triangle\phi|^{\mathfrak{U}}_{w,v} = 1$ iff, for all u such that vRu, $|\phi|^{\mathfrak{U}}_{u,v} = 1$.

(A_2) $|\triangle\phi|^{\mathfrak{U}}_{w,v} = 1$ iff, for all u such that wRu, $|\phi|^{\mathfrak{U}}_{u,v} = 1$.

Each of these proposals is unacceptable *qua* semantics. *Qua* model theory, each proposal, it will turn out, results in an unacceptable logic: each either validates some sentences that should not be validated or fails to validate some that should be. And in each case the logical flaw is closely related to a semantic flaw.

To begin, one serious logical shortcoming of (O_1) is that it fails to validate $\Box(\triangle\phi \supset \phi)$ — the principle that, necessarily, whatever is definitely so is so. By (O_1), $\Box(\triangle\phi \supset \phi)$ is false at a world w in a model just in case, for some world v of the model, $|\phi|_{w,v} = 0$, while, for all worlds u such that vRu, $|\phi|_{u,u} = 1$, and clearly there are models with worlds satisfying this condition.

Now a failure to validate something that should be validated does not immediately show that the model theory cannot deliver a correct semantics; for that it would also have to say some false things about the intended model. (Cf. if we let the interpretation of $\neg$ vary from model to model we would fail to validate every truth-functional tautology,[21] but this would not necessarily portend any

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

trouble with the semantics, since it would be consistent with ¬ being interpreted as negation in the intended model.) However, it is easy enough to see that the actual world of the intended model must witness the invalidity of $\Box(\triangle\phi \supset \phi)$ if (O_1) is correct, and this is why (O_1) is not acceptable for semantic purposes. For example, let ϕ be the sentence 'Hesperus ≠ Phosphorus.' In the actual world of the intended model — that is, actually — the proposition ϕ expresses is false in every world. However, in some other world, ϕ expresses a proposition that is true in every world. (Any world in which '≠' means what it actually means and 'Hesperus' and 'Phosphorus' refer to distinct planets in spite of their reference being fixed in the manner it actually is is such a world.) It follows that $\Box(\triangle\phi \supset \phi)$ is false. But $\Box(\triangle\phi \supset \phi)$ says that it is necessary that if it is definite that Hesperus ≠ Phosphorus then Hesperus ≠ Phosphorus, which is clearly true, so $\Box(\triangle\phi \supset \phi)$ is true, contrary to (O_1).

(O_2) also does not validate $\Box(\triangle\phi \supset \phi)$, and this portends trouble for its semantic applications. However, (O_2) has an even more serious problem that: it validates $\Box\triangle\phi \vee \Box\neg\triangle\phi$. This is immediately disqualifying, because validity entails truth. It is often a contingent matter what is definitely so. For example, I could have been definitely a pastry chef, and I could have been — because I am — definitely not a pastry chef.

Like (O_1), (A_1) fails to validate $\Box(\triangle\phi \supset \phi)$, and, as in the case of (O_1), this logical failure is accompanied by a semantic failure. By (A_1), $\Box(\triangle\phi \supset \phi))$ is false at any world w in any model with a world v such that $|\phi|_{w,v} = 0$, for all u such that vRu, $|\phi|_{u,v} = 1$, and clearly there are models satisfying this condition. Now consider the intended model. As in the discussion of (O_1), we can let ϕ be 'Hesperus ≠ Phosphorus.' Then, in the actual world of the intended model — i.e. actually — ϕ expresses a proposition that is false in every world. Thus, if there is also a world w such that, in every world close to w, ϕ expresses a proposition that is false in w, $\Box(\triangle\phi \supset \phi)$ is false. And, clearly that there is such a world. Consider the fact that, if 'Hesperus' and 'Phosphorus' had referred to different objects, ϕ would have expressed the same proposition — the necessary one — in every close world, and that proposition, being the necessary one, would have been true. Thus (since 'Hesperus' and 'Phosphorus' could have referred to different objects), there is a world w such that, in every world close to w, ϕ expresses a proposition that is true in w. Again, we reach the absurd conclusion that $\Box(\triangle\phi \supset \phi)$ — which says that it is necessary that if it is definite that Hesperus ≠ Phosphorus then Hesperus ≠ Phosphorus — is false.

Finally, while (A_2) does validate $\Box(\triangle\phi \supset \phi)$, it suffers from another logical flaw, which it shares with (A_1): it violates the principle I call *Definitization*, namely:

If $\vDash \phi$, then $\vDash \triangle\phi$.

Definitization says, in effect, that each valid sentence is (validly) definite. The models that are counterexamples to Definitization under (A_2) also turn out to

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

include the intended model, and this dooms (A_2) as a semantic proposal. $\phi \equiv A\phi$ is true at every proper point of every model, but, by (A_2), in any model,

$$\begin{aligned} |\triangle(\phi \equiv A\phi)|_w = 0 \quad &\text{iff} \quad \text{for some } v \text{ such that } wRv,\ |\phi \equiv A\phi|_{v,w} = 0 \\ &\text{iff} \quad \text{for some } v \text{ such that } wRv,\ |\phi|_{v,w} \neq |A\phi|_{v,w} \\ &\text{iff} \quad \text{for some } v \text{ such that } wRv,\ |\phi|_{v,w} \neq |\phi|_{v,v} \end{aligned}$$

For any sentence ϕ, then, $\nabla(\phi \equiv A\phi)$ is true iff ϕ expresses, in some world w R-related to the actual world, a proposition that has different truth values in w and in the actual world. And, it is reasonably clear that there are sentences that satisfy this condition, at least according to Williamson (1994): the worlds close to the actual world differ from the actual world with respect to various semantic states of affairs, and, provided that the semantic supervenes on the non-semantic, also with respect to various non-semantic states of affairs. If ϕ expresses some such state of affairs (proposition) both actually and in some close world, then $\nabla(\phi \equiv A\phi)$ is true. But this is absurd: we are in a position to know $\phi \equiv A\phi$ to be true, no matter what sentence ϕ might be, so, because borderline matters are unknowable, $\phi \equiv A\phi$ is not borderline.

In my view the above considerations are decisive against both of the otherworldly and both of the actualistic semantics for $\triangle$. There are, however, other ways of arguing against each option — ways that do not discriminate between (O_1) and (O_2), or between (A_1) and (A_2) — which it may be instructive to survey.

To begin with the otherworldly semantics, these share a flaw that can be leveraged to produce a variety of counterexamples. The shared flaw is the following. On any reasonable view of vagueness, not just Williamson's, it is not sufficient for it to be borderline whether S that it could have (easily or otherwise) not been the case that S. The actualistic semantics get this right; the otherworldly ones do not.[22]

One way to appreciate this flaw is to note that, according to the otherworldly semantics, a certain kind of metaphysical contingency is sufficient for borderlineness: if ϕ expresses the same proposition in all close worlds, and this proposition is contingent in such a way that it is true in some and false in other close worlds, then $\nabla\phi$ is true according to the otherworldly semantics. This is clearly wrong from the point of view of just about anyone's theory of vagueness. It is also wrong specifically from the point of view of Williamson's epistemicism, because of the latter's commitment to Semantic Plasticity, according to which a necessary condition for the truth of $\nabla\phi$ is that ϕ fails to express the same proposition in all close worlds. The otherworldly semantics allow $\nabla\phi$ to be true even when ϕ is not semantically plastic.

Let us next consider the actualistic semantics.

One problem with these is that they misclassify certain definite disquotational sentences as non-definite (both in the actual world and worlds close to the actual world). That they misclassify certain definite T-sentences as non-definite

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

is particularly damning, given that one of epistemicism's advertised virtues was its ability to respect the disquotational conception of truth.[23]

To present the problems that disquotational sentences pose for the actualistic semantics properly it will be necessary to extend the 2D model theory to deal with a language that results from adding some predicates and singular terms to L_{VM}. To that end, let a 2D model now be a quadruple $\langle W, R, D, [\![.]\!] \rangle$, where W and R are as before, D is a non-empty set representing the domain of individuals, and $[\![.]\!]$ is otherwise as before, except that, for all singular terms $\tau, \tau_1, \ldots, \tau_n$, and all n-place predicates ϕ,

(vi) $[\![\tau]\!]$ is a function from $W \times W$ to D

(vii) $[\![\phi]\!]$ is a function from $W \times W$ to $\mathcal{P}(D^n)$

(viii) $|\phi(\tau_1, \ldots, \tau_n)|^{\mathfrak{U}}_{w,v} = 1$ iff $\langle [\![\tau_1]\!](w, v), \ldots, [\![\tau_n]\!](w, v) \rangle \in [\![\phi]\!](w, v)$

In order to deal with disquotational sentences, we'll assume that, for each expression ξ, there is a singular term $\ulcorner\xi\urcorner$ (a quote name) such that, in the intended model,

(ix) If $@Rw$, then $[\![\ulcorner\xi\urcorner]\!](w, v) = \xi$

Now let us begin by considering *T-sentences*, i.e. sentences of the form $T(\ulcorner\phi\urcorner) \equiv \phi$, where T is the enriched language's truth predicate. For paradox-avoiding reasons we had better not assume that all T-sentences are true, so, since definiteness is factive, we had better not assume that all T-sentences are definite either. But let us say that a T-sentence is *healthy* when the sentence whose quote-name occurs in it is suitable in the sense of note 2. Certainly, we can safely assume that all healthy T-sentences are definite. And we can safely assume also all healthy T-sentences are definite in all close worlds, as used in those worlds. This is because a world is close to the actual world iff it is semantically indiscriminable from it; a world in which a healthy T-sentence, as used in that world, is not definite in that world, is one in which T-sentences are semantically different, and are semantically different in a discriminable way (since we know all healthy T-sentences are definite), from how they are in the actual world.

However, there is a straightforward argument for the conclusion that, if either of the actualistic semantics is correct, then some healthy T-sentence is not definite in some close world. The argument has two premises. The first is that, in the intended model,

(1) If ϕ is suitable, then $\phi \in [\![T]\!](@, w)$ iff $|\phi|_{w,w} = 1$.

(1) is, I take it, an uncontroversial minimal condition on the behavior of a truth predicate: it says that the truth predicate, as interpreted in the actual world, applies to a suitable sentence in a world w just in case that sentence, as used in w, expresses a proposition that is true in w. This follows from the trifling

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

observations that, first, necessarily, a suitable sentence is true iff the proposition it expresses is true, and, second, that, necessarily, a suitable sentence is true iff it belongs to the extension determined by the actual semantic value of 'true.'

(1), together with either of the actualistic semantics, implies:

(1*) If $@Rw$, $|\phi|_{@,w} \neq |\phi|_{w,w}$, and ϕ is suitable, then $|\triangle(T(\ulcorner\phi\urcorner) \equiv \phi)|_{w,w} = 0$.[24]

The second premise of the argument is:

(2) For some w and some suitable ϕ, $@Rw$ and $|\phi|_{@,w} \neq |\phi|_{w,w}$.

(1*) and (2) imply that there is a healthy T-sentence that is not definite in some close world, which is absurd.

The case for (1) is overwhelming. What about (2)? Note that anyone who accepts either actualistic semantics must accept, on pain of denying that there are any suitable borderline sentences, that there are worlds v, w, and a suitable sentence ϕ, such that vRw, $|\phi|_{v,w} \neq |\phi|_{w,w}$; otherwise, on either actualistic semantics, $\nabla\phi$ is false at every proper point of evaluation, so in particular is actually false. Given this, it would be very odd if there were no such cases in which $v = @$ — which is to say, if (2) were not true.

There is a similar argument to be made about disquotational *R-sentences* — sentences like

(E) 'Mount Everest' refers to Mount Everest.

To deal with these, we'll assume that the object language contains a reference predicate *Ref*. An R-sentence, then, has the form $Ref(\ulcorner\tau\urcorner, \tau)$. Clearly (at least, given our idealizations), we want to say that all R-sentences are definite. But there is a straightforward argument for the conclusion that this is not so if either of the actualistic semantics is correct.

This argument, too, has two premises. The first is that, in the intended model,

(3) $\langle x, y\rangle \in [\![Ref]\!](@, w)$ iff $[\![x]\!](w, w) = y$

(3) is minimal condition on *Ref* expressing the relation that 'refers' expresses in (E): (3) says, in effect, that, necessarily, a pair $\langle x, y\rangle$ belongs to the extension determined by the relation *Ref* is as *actually* used to express if and only if x refers to y. Since *Ref* is actually used to express the relation of reference, (3) reduces to the trivial observation that, necessarily, a pair $\langle x, y\rangle$ belongs to the extension determined by the relation of reference if and only if x refers to y.

(3), together with either of the actualistic semantics, implies:

(3*) If $[\![\tau]\!](w, w) \neq [\![\tau]\!](@, w)$ and $wR@$ then $|\triangle Ref(\ulcorner\tau\urcorner, \tau)|_{w,w} = 0$.[25]

The second premise of the argument is:

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

(4) For some singular term τ and some world w, $[\![\tau]\!](w, w) \neq [\![\tau]\!](@, w)$ and $wR@$.

(4) and (3*) entail that some R-sentence is non-definite in some close world, which is absurd.

The argument had two premises: (3) and (4). (3) was a truism. What about (4)? As in the case of the argument concerning T-sentences, it can be argued that the advocate of an actualistic semantics must accept (4) — in this case, on pain of denying the that there are borderline identity statements with proper names, given the natural way of introducing an identity predicate = into the semantics, which is via the clause

$$|\tau_1 = \tau_2|_{w,v} = 1 \text{ iff } |\tau_1|_{w,v} = |\tau_2|_{w,v}.$$

Assuming, again following Kripke (1980), that proper names are modally rigid, which is to say, in the 2D setting, that $|\tau|_{w,v} = |\tau|_{w,u}$, the only way for $\nabla(\tau_1 = \tau_2)$ to be true, on either actualistic semantics, is for there to be worlds w, v, such that $[\![\tau_1]\!](w, w) \neq [\![\tau_2]\!](v, w)$ and wRv. It would be very odd if there were no such case in which $v = @$.

(Further counterexamples involving other kinds of disquotational sentences — e.g. of the form '$\forall x$('F' applies to $x \equiv F(x)$)' — can be devised, but I'll leave them as an exercise for the reader.)

The problems for the actualistic semantics do not end there. I'll mention one further problem that will make an interesting test case for further semantic proposals. Consider the *Meter Sentence* made famous by Kripke (1980, 54f):

> The Standard Meter is one meter long.

The Standard Meter is, of course, the object whose length is used to fix the reference (and content) of 'meter.' Since the length of the Standard Meter — a platinum bar cast using nineteenth-century technology — could very easily have been different from its actual length, the Meter Sentence could very easily have expressed a proposition that is actually false. It is also highly plausible that some of the worlds in which the Meter Sentence expresses a proposition that is actually false, are worlds that are close to the actual world in the sense of 'close' that matters to the epistemicist interpretation of $\triangle$; but if so, it follows by either actualistic semantics that the Meter Sentence is borderline, so, because borderline status is incompatible with knowledge, no one knows that the Standard Meter is one meter long, which is absurd.

(Counterexamples like the above can be multiplied indefinitely. 'Kilogram,' 'second,' and other measure words whose reference is fixed by paradigm objects whose relevant measurements could very easily have been different than they actually are easily lend themselves to similar counterexamples. So, in general

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

do cases in which we use a reference-fixing description that could very easily have had a referent different from its actual referent.)

I think the arguments given so far establish the incorrectness of both of the otherworldly semantics and both of the actualistic semantics, but they do not come close to showing that there is *no* 2D semantics for L_{VM} acceptable to epistemicists. After all, I have only considered four possible truth definition clauses for $\triangle$ within 2D model theory — the ones corresponding to the two proposals Caie (2012) finds the most 'natural' — but this still leaves a large number of alternative truth definition clauses unexamined.[26] Are we in a position to know that all of the alternatives are also unacceptable?

In fact, we are, at least if we impose some further natural conditions on the adequacy of an epistemicist semantics for L_{VM}. We only need two conditions in addition to Definitization. The first condition is that every instance of the schema

K: $\quad \triangle(\phi \supset \psi) \supset (\triangle\phi \supset \triangle\psi)$

should be true in the intended model. The second is that there should be no vagueness in the world, in a sense that will be made precise below. A result by Peter Fritz, which is included in the Appendix to this paper (Fritz 2016), shows that there is a precise sense in which there is no 2D semantics for L_{VM} that satisfies all three conditions.

The idea that there is no vagueness in the world is most naturally expressed using propositional quantification, although we will later see that this is dispensable. Let us, then, begin by considering the language L_{VMQ}, which results from the addition of propositional variables and the universal quantifier $\forall$ to L_{VM}, with the usual formation rules. In L_{VMQ}, we express the claim that there is no vagueness in the world by the sentence

NVW: $\quad \forall p(p \equiv \triangle p)$,

which says, in effect, that every state of affairs (proposition) obtains iff it definitely obtains. Semantic Plasticity mandates NVW: if every borderline formula could easily have had expressed a proposition other than the proposition it actually expresses, a formula that is a variable cannot be borderline because, under any given variable assignment, it could not have expressed any proposition other than the one it actually expresses. The same goes, *mutatis mutandis*, for variables of all types. (For similar reasons, one would expect a supervaluationist to endorse NVW.)[27]

Now of course an epistemicist who endorses NVW must reject the validity of universal instantiation on propositional variables, as that rule would take us from the plausible NVW to each instance of the absurd schema

TRIV: $\quad \phi \equiv \triangle\phi$,

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

which says, in effect, that there is no vagueness. But there is nothing new here: it is natural for epistemicists (and indeed supervaluationists) to reject the validity of universal instantiation on variables of all types. For example, in a first-order language with $\triangle$, it is natural for an epistemicist (or supervaluationist) to accept $\forall x \forall y(x = y \supset \triangle x = y)$ while rejecting some instances of $a = b \supset \triangle a = b$, because there are vague names.[28]

Let us next consider how we should augment the 2D model theory in order to handle the L_{VMQ}. Intuitively, the propositional variables range over propositions. In the model theory, however, we must interpret them, under a variable assignment, as entities of the same type as a model's interpretation function assigns to the atomic sentences, which is to say, as sets of points of evaluation of the model. Now clearly, the admissible values of the propositional variables cannot be *all* sets of such pairs: some of them represent context-dependent as well as circumstance-dependent variation in truth value, whereas the truth value of a proposition depends only on the circumstance. The admissible values of the propositional variables, then, must satisfy the condition

$$\langle w, v\rangle \in X \qquad \text{iff} \qquad \langle u, v\rangle \in X.$$

Let's say that any set of points of evaluation satisfying the above condition is a *barcode*.[29] In the intended model, the subsets of $W \times W$ that are barcodes, then, represent the propositions. We'll define a *variable assignment* on a model as any function from the propositional variables to the model's barcodes. Now the notion of the truth value $|\phi|^{\mathfrak{U}^g}_{w,v}$ of a formula ϕ at a point of evaluation $\langle w, v\rangle$ in a model $\mathfrak{U}$ is relativized to a variable assignment g. $|\phi|^{\mathfrak{U}^g}_{w,v}$ will be defined exactly like $|\phi|^{\mathfrak{U}}_{w,v}$ except for when ϕ is a propositional variable or a universally quantified formula; for these cases we add (xii) and (xiii).

(xii) $|p|^{\mathfrak{U}^g}_{w,v} = 1$ iff $\langle w, v\rangle \in g(p)$

(xiii) $|\forall p\,\phi|^{\mathfrak{U}^g}_{w,v} = 1$ iff, for each variable assignment f on $\mathfrak{U}$ differing from g at most in what f assigns to p, $|\phi|^{\mathfrak{U}^f}_{w,v} = 1$

Logical consequence is defined as truth preservation at each proper point of each model under each variable assignment on that model.

Now it turns out that, no matter what truth definition clause we introduce for $\triangle$ in L_{VMQ}, if the clauses already introduced are in place and the model theory validates every instance of *K* and satisfies Definitization, every model in which NVW is true at every proper point will also be a model in which every instance of TRIV is true at every proper point. So, in particular, this will hold of the intended model, and there is no hope of getting an acceptable epistemicist semantics for L_{VMQ} out of a 2D model theory. This is a consequence of Fritz's Proposition 1, which is more general. Proposition 1 states that there is no way of interpreting $\triangle$ by a Scott–Montague neighborhood function that satisfies our desiderata

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

without validating every instance of TRIV. Each way of interpreting $\triangle$ in a relational model-theoretic semantics of the kind I have been working with in this section corresponds to a neighborhood function, but not conversely. There is, then, a good sense in which Fritz's result shows that there is no 2D model theory that can serve as the basis of an acceptable epistemicist semantics for L_{VMQ}.

Fritz's result does not rely essentially on the presence of propositional quantification in the language. While we cannot express the claim that there is no vagueness in the world in L_{VM}, we can express that claim in the 2D semantics of L_{VM} by requiring that, for each barcode b, the neighborhood function that interprets $\triangle$, in effect, classifies b as definite at a proper point of evaluation iff b is true at that point. Let us call this condition *NVW**. If we assume that the interpretation of $\triangle$ obeys NVW*, and we assume Definitization and the validity of each instance of K, it once again follows that each instance of TRIV is true. For let us say, following Fritz, that a *generalized 2D frame* is a pair $\langle W, D\rangle$, with W as before and D a neighborhood function that interprets $\triangle$, and let us say that the *quantifier-free logic* of a class of generalized 2D frames is the set of all L_{VM}-sentences true at every proper point of every model based on a frame in that class. Fritz's second result (Proposition 2) states that the logic of any class of generalized 2D frames satisfying NVW* that includes all instances of *K* and is closed under the rule $\phi/\triangle\phi$ (corresponding to Definitization) also includes all instances of TRIV. Now if we make the further natural assumptions that there is an intended generalized 2D frame, on which the intended model is based, and whose logic includes each instance of *K* and is closed under the rule $\phi/\triangle\phi$, it follows that if this frame satisfies NVW*, then each instance of TRIV is true at each proper point of the intended model, so is true *simpliciter*.

5. The failure of the 2D approach to the semantics of L_{VM} speaks in favor of a move to a 3D approach. The model theory turns out to be straightforward, and it has already proved to be fruitful in logical investigations: Litland and Yli-Vakkuri (2016) show how their 3D model theory can be used to obtain completeness results for certain propositional logics of vagueness and modality, at least one of which is plausibly the correct logic from an epistemicist point of view. However, what semantics the epistemicist should marry with this model theory is as yet unclear. In this section I will introduce a 3D model theory for L_{VM} that is a simplified version of one of the model theories discussed by Litland and Yli-Vakkuri, and I will ask how we might go about extracting an epistemicist semantics from it.

Let us say that a *3D model* is a triple $\mathfrak{U} = \langle W, R, [\![.]\!]\rangle$, where W is a non-empty set, R a reflexive and symmetric relation on W, and $[\![.]\!]$ a function from the atomic sentences to 3D points of evaluation in $\mathfrak{U}$, which are subsets of $W \times W \times W$.

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

The truth value $|\phi|^{\mathfrak{U}}_{w,v,u} \in \{0, 1\}$ of a sentence ϕ at $\langle w, v, u\rangle \in W \times W \times W$ in $\mathfrak{U}$ is defined as follows, where α is an atomic sentence and ϕ and ψ are any sentences.

(i)	$\lvert\alpha\rvert^{\mathfrak{U}}_{w,v,u} = 1$	iff	$\langle w, v, u\rangle \in [\![\alpha]\!]$
(ii)	$\lvert\neg\phi\rvert^{\mathfrak{U}}_{w,v,u} = 1$	iff	$\lvert\phi\rvert^{\mathfrak{U}}_{w,v,u} = 0$
(iii)	$\lvert\phi\wedge\psi\rvert^{\mathfrak{U}}_{w,v,u} = 1$	iff	both $\lvert\phi\rvert^{\mathfrak{U}}_{w,v,u} = 1$ and $\lvert\psi\rvert^{\mathfrak{U}}_{w,v,u} = 1$
(iv)	$\lvert\Box\phi\rvert^{\mathfrak{U}}_{w,v,u} = 1$	iff	for all $u' \in W$, $\lvert\phi\rvert^{\mathfrak{U}}_{w,v,u'} = 1$
(v)	$\lvert A\phi\rvert^{\mathfrak{U}}_{w,v,u} = 1$	iff	$\lvert\phi\rvert^{\mathfrak{U}}_{w,v,u'} = 1$
(vi)	$\lvert\triangle\phi\rvert^{\mathfrak{U}}_{w,v,u} = 1$	iff	for all w' such that wRw', $\lvert\phi\rvert^{\mathfrak{U}}_{w',v,u} = 1$

Logical consequence is defined just like before, except in that one of the terms occurring in its definition is re-refined: we now say that a proper point of evaluation is a point of the form $\langle w, w, w\rangle$. We will also say that ϕ is true in w iff ϕ is true at the proper point of w, i.e. at $\langle w, w, w\rangle$.

As the reader can verify, this model theory does not share any of the logical flaws of the otherworldly and actualistic 2D model theories discussed in Section 4. In fact, there is some plausibility to the idea that the relation of logical consequence that it — or some close variant of it[30] — delivers is the correct logic of vagueness and modality from an epistemicist point of view. I won't press the case for that here, however. I do so in my joint work with Litland, which discusses both epistemicist and supervaluationist applications of the 3D model-theoretic framework.[31]

It is also worth noting that the 3D model theory can easily be extended to handle L_{VMQ} from Section 4. The extension is carried out using exactly the same definition of the notion of truth at a point in a model under a variable assignment. Only the definition of a barcode will be different: in the 3D setting a barcode of a model is any set X of points of evaluation of the model such that $\langle w, v, u\rangle \in X$ iff $\langle w', v', u\rangle \in X$. This model theory validates NVW, as desired, but it does not validate every instance of TRIV.

So far we only have a 3D model theory for L_{VM}, but what we want is a way of extracting a semantics for L_{VM} from that model theory. This requires us to make sense of the idea of a sentence being true at a triple of worlds. I will discuss two ways of doing so.

I will call first way of extracting a semantics for L_{VM} from the 3D model theory the *metasemantic interpreration* of that model theory. According to the metasemantic interpretation, ϕ is true at $\langle w, v, u\rangle$ iff the *character* ϕ has in w, when applied to the *context* v yields a proposition that is true in the *circumstance* u. R can be, as before, glossed as a relation of 'semantic indiscriminability,' except in this case the semantic facts with respect to which the worlds it relates are indiscriminable are facts about the characters rather than about the (propositional) contents of sentences.[32] The metasemantic interpretation recognizes

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

two different roles for contexts, which I'll call the role of *metasemantic context* (the role of the first world in a triple) and simply *context* (the role of the second world in a triple). The metasemantic context represents the dependence the characters of L_{VM} sentences on global patterns of language use, and generally on whatever else facts about character supervene on. The context is just a context in Kaplan's (1977) sense: once the character of a sentence is fixed by a metasemantic context, that character is applied to a context to obtain a content. According to the metasemantic interpretation, then, since the character of a sentence is its actual character, the character of ϕ is the function $\lambda w \lambda v |\phi|_{@,w,v}$. A suitable characterization of the function $\lambda\phi\lambda w\lambda v|\phi|_{@,w,v}$ will be a semantics.

If the metasemantic interpretation is correct, then the kind of 'semantic value' mentioned in the statement of Semantic Plasticity must be character rather than (propositional) content. It is a consequence of the metasemantic interpretation that, if $\nabla\phi$ is true in w, then for some v close to w, the character of ϕ in v is different from the character of ϕ in w. But it is not a consequence of the metasemantic interpretation that, if $\nabla\phi$ is true in w, then for some v close to w, the content of ϕ in v is different from the content of ϕ in w.

Suppose that the metasemantic interpretation is correct. Then it will be tempting to think that there is a straightforward account of why it gets right all the cases that made trouble for the 2D actualistic semantics from Section 4.

Consider first the case of $\triangle(\phi \equiv A\phi)$. Here, the 3D model theory alone solves the problem faced by the actualistic 2D model theories: because the 3D model theory validates each instance of $\phi \equiv A\phi$ and satisfies Definitization, it also validates each instance of $\triangle(\phi \equiv A\phi)$. A valid sentence is true at every proper point of the intended model, so is true *simpliciter*. So, each instance of $\triangle(\phi \equiv A\phi)$ is true.

The metasemantic interpretation, however, also delivers an explanation of *why* the model theory gets the case of $\triangle(\phi \equiv A\phi)$ right, as follows.

Let's say that a character f is *diagonally true* iff, for all w, $f(w)(w) = 1$. As we have learned from Kaplan, the character of any instance of $\phi \equiv A\phi$ is diagonally true. That the character of each instance of $\phi \equiv A\phi$ is diagonally true is a semantic fact about instances of $\phi \equiv A\phi$, in the sense of 'semantic' relevant to semantic discriminability under the metasemantic interpretation. It is also a known semantic fact. But since it is, in the relevant sense, a known semantic fact that the character of each instance of $\phi \equiv A\phi$ is diagonally true, then each instance of $\phi \equiv A\phi$ has a diagonally true character in every metasemantic context semantically indiscriminable from (R-related to) the actual one. It immediately follows that, in every metasemantic context R-related to the actual one, each instance of $\phi \equiv A\phi$ has a character that, when applied to the actual world, gives a proposition that is true in the actual world, and this, by the metasemantic interpretation and (vi), is equivalent to saying that each instance of $\triangle(\phi \equiv A\phi)$ is true.

It is tempting to generalize this explanation to all of the problematic cases, and to conjecture that T-sentences, R-sentences, the Meter Sentence, etc., all have characters that are known to be diagonally true (both in the actual

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

metasemantic context and in ones *R*-related to it), so the definitizations of those sentences are true (both in the actual metasemantic context and in ones close to it). Here is why that line of thought is tempting.

Consider, to begin, either kind of disquotational sentence. Even though the propositions expressed by, e.g. '"Tim is thin" is true iff Tim is thin' and '"Tim" refers to Tim' are contingent, each sentence appears to be associated with a kind of semantic guarantee of truth. One natural way to make precise the idea of a sentence enjoying a semantic guarantee of truth is to say that the sentence has a diagonally true character: a diagonally true character guarantees that, in any world, the sentence expresses a proposition that — even though it may be a contingent proposition — is true in that world. Similarly, even though the Meter Sentence expresses a contingent proposition, and one that could very easily have been false, there is something about the convention we use to fix the content of 'meter' that guarantees that the sentence cannot be asserted falsely (as long as that convention is in effect, and as long as the other words in the sentence have their actual conventional meanings). If we take the further step of assuming that the convention in question is the character of 'meter,' then it will be tempting to conclude that the character of the Meter Sentence is also diagonally true. And similarly for other similar sentences involving 'kilogram,' 'second,' etc. If the characters of all of the problematic sentences are diagonally true in all worlds close to the actual one, and we are in a position to know this in all worlds close to the actual one, then they will all be definite, according to the metasemantic interpretation of the 3D model theory, in all worlds close to the actual one.

Superficially, at least, this line of thought seems attractive. Yet it is, I think, quite clearly incorrect. The metasemantic interpretation faces two serious problems, the second of which I find to be an especially decisive consideration against adopting it.

The first problem is this. Under the metasemantic interpretation, the 3D model theory only differs from the otherworldly 2D semantics on whether $\triangle\phi$ is true at the proper point of w (i.e. at $\langle w, w\rangle$ according to the 2D model theory and at $\langle w, w, w\rangle$ according to the 3D model theory) when ϕ has a non-constant character (i.e. is an indexical sentence) in some world close to w. It follows that disquotational sentences, the Meter Sentence, etc., must be indexical, in the actual world or worlds close to it, in subtle ways we haven't noticed before. Since a sentence is indexical in a world if and only if at least one of its simple constituents is,[33] these sentences must contain indexical words: apparently 'true,' 'refers,' 'meter,' and so on. Yet it is very unclear what the non-constant characters of the semantic predicates might be. In the case of 'meter,' one might think, as suggested above, that the non-constant character is the function that assigns to each world w the property of being as long as the Standard Meter is in w (or the content that represents that property). But one might also think, more plausibly, that this line of thought confuses semantic with metasemantic matters:

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

the convention that fixes the content of 'meter' also fixes its character, which is constant. Thus, if the length of the Standard Meter had been different from its actual length, the Meter Sentence would have had both a different character and a different content. And indeed this is the consensus on how so-called 'reference-fixing descriptions' work: they do not introduce indexicals, but rather expressions whose character as well as content depends on what satisfies the description.[34] If the consensus is correct, as I take it to be, then the metasemantic interpretation is incorrect.

(What about the above sketch of an argument for the conclusion that disquotational sentences and the Meter Sentence have diagonally true characters in all worlds close to the actual world? It confuses two properties: that of having a diagonally true character and that of having a character that determines a true proposition. It is plausible that each of the problematic sentences has, in each world *R*-related to the actual one, a character that, when applied to *w*, yields a proposition that is true in *w*. But it does not follow from this — and nothing has been said to support the claim — that each of the problematic sentences has, in each world *R*-related to the actual one, a character that, when applied to an arbitrary world *v*, yields a proposition that is true in *v* — i.e. a diagonally true character.)

The second problem with the metasemantic interpretation was noticed by John Hawthorne. In conversation, Hawthorne observed that we can refer to the actual world by both indexical and non-indexical means, and he used this to make trouble for the metasemantic interpretation. Here, I will present a simplified version of Hawthorne's objection:[35]

The indexical means of referring to referring to the actual world in L_{VM} is, of course, to use *A*. But we could also introduce a non-indexical 'actuality' operator into a language as follows. First, we introduce a non-indexical proper name, say, 'Worldy,' for the actual world into our metalanguage; then we enrich L_{VM} with an operator A^*, such that $A^*\phi$ translates into the metalanguage as 'In Worldy, it is the case that ϕ.' Since we know that A^* has a constant character,[36] A^* must, by familiar reasoning, have a constant character in every world *R*-related to the actual one. In particular, it must be that:

(1) If $@Rw$, then $|A^*\phi|_{w,v,u} = 1$ iff $|\phi|_{w,w,w} = 1$.

It follows that each instance of $A^*\phi \equiv A\phi$ is true, and, having noticed this, we know that each instance of $A^*\phi \equiv A\phi$ is true. But from (1) and the rest of our semantics it follows that an instance of $\triangle(A^*\phi \equiv A\phi)$ is false whenever $|\phi|_{w,w,w} \neq |\phi|_{w,@,@}$, for some *w* such that $@Rw$. That there is a world *w* and a sentence ϕ such that $|\phi|_{w,w,w} \neq |\phi|_{w,@,@}$ is pretty clear. (According to the semantics, there is vagueness only if there are *R*-related worlds *w*, *v* such that $|\phi|_{w,w,w} \neq |\phi|_{w,v,v}$, and it would be odd if there were no such case in which $v = @$.) But then such an instance of $A^*\phi \equiv A\phi$ will be a counterexample to the principle that knowledge entails definiteness.

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

This latter problem, which I take to be decisive, exposes a weakness that the metasemantic interpretation of the 3D model theory shares with both of the actualistic 2D semantics. According to all of these semantics it is sufficient for the truth of $\nabla\phi$ at a proper point that ϕ is false at an *improper* point suitably related to it. But no one ever uses, or could use, a sentence at an improper point, so, *a fortiori,* no one ever uses, or could use, a sentence to say something false at an improper point. Falsehood at such a point is not a genuine possibility of error, so is not the kind of possibility of error whose closeness precludes knowledge.

The second way of extracting a semantics from the 3D model theory, which I find to be more promising, is the *epistemic interpretation*. According to the epistemic interpretation, the first coordinate of a 3D point of evaluation is not a metaphysically possible world, but an epistemic possibility, in a certain broad sense, about which I will have more to say below. I will use '*e*,' '*e*′,' etc., as variables for these (broadly speaking) epistemic possibilities, and I will call the epistemic possibility that obtains in *w* '*e*(*w*)'; the actual epistemic possibility, then, is designated by '*e*(@)'. The rough idea is that ϕ is true at $\langle e, w, v\rangle$ iff the proposition ϕ expresses in *w*, according to *e*, is true in *v*. A proper point of evaluation will be of the form $\langle e(w), w, w\rangle$. (Thus 'epistemic interpretation' turns out to be a bit of a misnomer: it is not an merely an interpretation of the 3D formal apparatus, since it requires our semantics for L_{VM} to draw finer distinctions than that apparatus does, such as the distinction between *e*(*w*) and *w*. I will return to the importance of keeping epistemic possibilities and metaphysically possible worlds separate below.)

The epistemic interpretation makes no use of the Kaplanian distinction between character and content. When an epistemic possibility *e* represents a sentence ϕ as expressing different propositions in worlds *w* and *v*, this could come about in two ways: either ϕ has different constant characters in *w* and in *v* according to *e* or ϕ has the same character in *w* and in *v* according to *e*, but this character is variable and determines different contents in *w* and in *v*. The epistemic interpretation is entirely insensitive to whether a given difference in content is due to a difference in character. For this reason it is not vulnerable to the first problem faced by the metasemantic interpretation: under the epistemic interpretation, that the 3D semantics and the 2D actualistic semantics differ on whether a sentence is definite at a proper point of evaluation tells us nothing about whether the sentence is indexical at any point.

According to the epistemic interpretation, the function $\lambda\phi\lambda w\lambda v|\phi|_{e,w,v}$ takes each sentence to the function that takes each world to the content that sentence expresses in that world, according to the epistemic possibility *e*. Which proposition a sentence expresses in a world is not a contingent matter; therefore (by (vi) and the fact that there is vagueness) some of the epistemic possibilities are metaphysically impossible. *e* will be metaphysically possible only if $\lambda\phi\lambda w\lambda v|\phi|_{e,w,v} = \lambda\phi\lambda w\lambda v|\phi|_{e(@),w,v}$. The only feature of *e* relevant to the envisioned semantics is the way in which content facts supervene some of the facts

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

according to *e*, which is represented by the function $\lambda\phi\lambda w\lambda v|\phi|_{e,w,v}$. The bare-bones version of the epistemic interpretation offered here is neutral on the question of just which facts constitute the supervenience base in question. One answer, which is in the spirit of Hawthorne (2006), is that they are all of the 'metaphysical groundfloor' facts in some sense — perhaps the microphysical facts. Another possible answer is that they are all of the facts about the use of language, where 'use' is construed broadly enough to guarantee that no two worlds differ with respect to content facts without differing with respect to the use facts. However the details are filled in, the epistemic interpretation is committed to ignorance due to vagueness being explained by a kind of principled ignorance of the way in which semantic (content) facts supervene on certain other facts. In this respect, it represents a departure from the epistemicism of Williamson (1994): it does not include a commitment to Semantic Plasticity (although it is consistent with it) — a matter to which I'll return below.

For this reason, on the epistemic interpretation, *R* can no longer be glossed as 'semantic indiscriminability.' Rather, points of the first dimension that are *R*-related to the actual point *e*(@) represent ways in which, for all we (or creatures relevantly like us) are actually in a position to know, the content facts supervene on certain other facts. The points *R*-related to points *R*-related to *e*(@) represent ways in which, for all we are actually in a position to know, for all we are in a position to know, the content facts supervene on certain other facts, etc. For this reason I called the points of the first dimension epistemic possibilities 'in a broad sense.' Assuming that all of the points are related to *e*(@) by the ancestral of *R*, strictly speaking the only epistemic possibilities are those *R*-related to *e*(@), while the others are either merely epistemically possibly epistemically possible, or merely epistemically possibly epistemically possibly epistemically possible, etc. Nevertheless, for lack of a better term, I will continue to call these points 'epistemic possibilities.'

Now, it is fairly clear that the epistemic interpretation does not have a problem accounting for the definiteness of disquotational sentences, the Meter Sentence, or other similar cases involving reference-fixing convention. For consider the version of the epistemic interpretation according to which the points on the first dimension represent epistemic possibilities for the content facts to supervene on the use facts. It is highly plausible that, while I am largely in the dark about the details of the supervenience function for, say, the Meter Sentence, I am in a position to know enough about that function to rule out its being one that assigns to the actual use facts a proposition that is false in the actual world.[37] And the same goes for disquotational sentences and for the other problematic cases involving reference-fixing conventions, as well as, indeed, for the Hawthorne-inspired example, $A^*\phi \equiv A\phi$.

Now one might worry that the epistemic interpretation nevertheless fails to address Hawthorne's problem: after all, what's to stop someone from adding to L_{VM} an operator $A^\dagger$ for which it is simply stipulated that

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

(1′) If $e(@)Re$, then $|A^{\dagger}\phi|_{e,v,u} = 1$ iff $|\phi|_{e,e,e} = 1$?

Then, by familiar reasoning, some instances of $A^{\dagger}\phi \equiv A\phi$ will be borderline but known to be true. The answer to this worry is simply that, according to the kind of semantics under consideration, $|\phi|_{e,e,e}$ is undefined, because $|\phi|_{e,w,v}$ is defined only when w and v are metaphysical, not epistemic, possibilities. (This is where the 3D model theory will require some tweaking if it is to deliver the intended model.)

One might be inspired by Hawthorne's example to have the following further worry. If we introduce an operator $A^{\ddagger}$ for which the condition

(1″) $|A^{\ddagger}\phi|_{e,v,u} = 1$ iff $|\phi|_{e(@),v,u} = 1$

holds, and we know that (1″) holds then, whenever ϕ is borderline, $\phi \equiv A^{\ddagger}\phi$ will also be borderline, yet known to be true.

I am not entirely clear on what the correct answer to this worry is, but, whatever it is, it is not particularly incumbent upon the advocate of the epistemic interpretation to come up with it. An 'epistemic actuality' operator like $A^{\ddagger}$ would wreak havoc on the semantic applications of any standard model theory for epistemic logic.[38] For suppose that there is a knowledge operator K such that, as is standard, we treat $K\phi$ as true in an epistemic possibility e iff ϕ is true in every epistemic possibility relevantly accessible (by a reflexive relation) from e. Then, if there is a unique actual epistemic possibility $e(@)$ such that ϕ is true in $e(@)$ iff ϕ is true, and there is an operator $E^{@}$ such that $E^{@}\phi$ is true in an arbitrary epistemic possibility e iff ϕ is true in $e(@)$, then every instance of $K(\phi \equiv E^{@}\phi) \supset (K\phi \vee K\neg\phi)$ is true. But this is absurd: if $E^{@}$ is interpretable in this manner at all, then it will often be much more difficult to know whether ϕ is true than to know that $\phi \equiv E^{@}\phi$ is true. We can come to know the latter simply by coming to know the semantic stipulation by which $E^{@}$ was introduced plus disquotation, but we cannot come to know all truths in this way. Since, according to the epistemic interpretation, the first coordinate of a point of evaluation is an epistemic rather than a metaphysical possibility, the envisaged introduction of $A^{\ddagger}$ by (1″) amounts to the introduction of an 'epistemic actuality' operator, which is presumably impossible for whatever reason the introduction of $E^{@}$, as described in this paragraph, is impossible. And if it is not impossible, then this points to a crisis in the foundations of the standard approach to the model theory of languages with epistemic operators, rather than a problem for the epistemic interpretation of the 3D model theory for L_{VM} in particular.

Finally, I will mention, without attempting to answer, two open questions about the epistemic interpretation. The first concerns Semantic Plasticity. The epistemic interpretation notably differs from the metasemantic interpretation in that the former does not entail Semantic Plasticity, for any natural interpretation of 'semantic value.' It does entail, of course, that when ϕ is borderline, ϕ expresses in the actual world, according to some close epistemic possibility, a

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

proposition other than the one it actually expresses in the actual world. But, as we have already seen, any such epistemic possibility will be metaphysically impossible, because it is a non-contingent matter which proposition is actually expressed by a sentence. Semantic Plasticity requires that a borderline sentence *could have* — in an alethic rather than epistemic modal sense — easily had a semantic value other than its actual semantic value. The epistemic interpretation is consistent with Semantic Plasticity: we could consistently add to it the claim that whenever a sentence ϕ expresses in the actual world, according to a close epistemic possibility, some proposition *p* other than the one it actually expresses in the actual world, there is also a metaphysical possibility that could easily have obtained (in which the use facts are ever so slightly different from the actual use facts) in which ϕ does express *p*. But so far it is unclear why the existence of such a metaphysical possibility would be required for an explanation of ignorance of borderline matters, since the epistemic interpretation already makes available an alternative explanation in terms of some kind of principled ignorance of the supervenience of content facts on certain other facts.

At least one issue in this area seems reasonably clear, which is that an explanation of ignorance of borderline matters in terms of Semantic Plasticity cannot require all of the close possibilities of error that preclude knowledge of borderline matters to be metaphysical possibilities. That is, it cannot be that, whenever ϕ is borderline, there is a close metaphysically possible world *w* such that: in *w* ϕ expresses some proposition *p* other than the proposition ϕ actually expresses, and the truth value *p* has in *w* is different from the actual truth value of the proposition ϕ actually expresses.[39] To see why this cannot be so, note first that the proposed principle entails (*).

(*) Whenever ϕ is true and borderline, there is a close metaphysically possible world in which ϕ expresses a proposition other than the proposition ϕ actually expresses, and in which that proposition is false.

We can use Semantic Plasticity to argue against (*) as follows. Consider the case of a non-semantically plastic, so, by Semantic Plasticity, non-borderline sentence $\sharp$ that specifies the all of the actual use facts. Let χ be any true borderline sentence, and let $p_{@}$ be the proposition actually expressed by χ. Then, given the supervenience of content facts on use facts,

(**) It is necessary that, if the proposition actually expressed by $\sharp$ is true, then χ expresses $p_{@}$.

Since χ is borderline but $\sharp$ is not, it follows, on very minimal assumptions about the logic of definiteness,[40] that $\sharp \supset \chi$ is borderline. $\sharp \supset \chi$ is also true, because χ is true. By (*), then, there is a close metaphysically possible world in which $\sharp \supset \chi$ expresses a proposition other than the proposition $\sharp \supset \chi$ actually expresses, and

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

in which that proposition is false. To see that there is no such world, suppose for a contradiction that there is one — call it '*w**' — and consider two cases:

First case: the proposition ♯ actually expresses is true in *w**. It follows by (**) that χ expresses $p_{@}$ — i.e. the proposition χ actually expresses — in *w**. Since ♯ is not semantically plastic, ♯ expresses what it actually expresses in *w** as well. Consequently, $\sharp \supset \chi$ expresses what it actually expresses in *w**, contrary to hypothesis.

Second case: the proposition ♯ actually expresses is not true in *w**. Again, since ♯ is not semantically plastic, ♯ expresses what it actually expresses in *w**. It follows that the proposition ♯ expresses in *w** is not true in *w**, and so it also follows that the proposition ♯ expresses in *w** is false in *w**, and that the proposition $\sharp \supset \chi$ expresses in *w** is true in *w**, contrary to hypothesis.[41,42]

The second open question concerns the principled ignorance of the supervenience of content on use that the advocate of the epistemic interpretation must posit.[43] I have no explanation of it to offer, but I will note, following Williamson, that any epistemicist must posit unknowable necessary truths:[44] e.g. if a net worth of $5,000,000 is the cut-off for wealth, then it is a necessary but borderline, and therefore unknowable, truth that to be wealthy is to have a net worth of at least $5,000,000. Of course, Williamson offers an explanation of the unknowability of such truths in terms of Semantic Plasticity, but the above argument casts doubt on the idea that such an explanation will be available in every case of ignorance of borderline matters.

6. I have argued that an acceptable epistemicist semantics for a language containing a definiteness operator $\triangle$ along with operators for metaphysical necessity ($\Box$) and metaphysical actuality (*A*) cannot be developed using the resources of the kind of 2D model theory that is standardly used in logical and semantic investigations of the interaction of $\Box$ and *A* alone. This suggests a move to the 3D model-theoretic framework that I developed with Jon Litland. The latter has already proved to be fruitful for logical investigations, but it remains an open question how, if at all, it can serve as the basis for a satisfactory epistemicist semantics for the interaction of $\triangle$, $\Box$, and *A*. I have sketched two ways in which one might attempt to extract an epistemicist semantics from the 3D model theory. The first of these was the metasemantic interpretation of the model theory, which vindicated a form of Semantic Plasticity, but with Kaplanian characters rather than contents (propositions) in the role of semantic value. The second of these was the epistemic interpretation, according to which the first dimension of semantic evaluation represents, in a certain broad sense, all of the epistemically possible ways for the content facts to supervene on certain other facts. While enjoying some superficial plausibility, the metasemantic interpretation turned out to have implausible consequences concerning indexicals. My own money is on the epistemic interpretation, or something close to it, delivering the correct epistemicist semantics, but if it or something like it does so, it is an interesting open question what role is left for Semantic Plasticity to play in the explanation of our ignorance of borderline matters.

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

Notes

1. Hawthorne (2006).
2. That is to say, a notion of truth that satisfies each instance of the schema '"*S*" is true iff *S*,' where instances are obtained by replacing '*S*' with a suitable sentence. A sentence is suitable just in case inserting it into the schema does not result in any trouble, such as, but not limited to, inconsistency (e.g. sentences that purport to ascribe truth to themselves are also not suitable). The relevant notion of suitability is notoriously difficult to make precise.
3. Williamson (2003, 710).
4. An equivalent definition of definiteness in terms of borderline cases is perhaps more intuitive (it is definite that *S* iff *S* and it is not borderline whether *S*), but in this paper I take definiteness as primitive, as this is more natural for logical and semantic investigations, in which 'definitely' is standardly treated like 'necessarily.'
5. For example, Shapiro (2007) has this to say about epistemicism: 'Here I do not muster a sustained argument against that view, and it is not polite to stare' (7). This remark is directed at both Williamson (1994) and Sorensen's (1988) earlier development of an epistemic theory of vagueness.
6. See Hawthorne (2006), Kearns and Magidor (2008), Caie (2012), and Magidor (forthcoming).
7. See Williamson (1994, Section 8) and (2003, 710).
8. The content of *A*, as of any indexical, will vary with context. In this paper, the logical constants are given a syncategorematic treatment — the usual practice — but they could also be assigned characters in obvious ways.
9. A different approach is suggested by a passage in Kaplan's 'Demonstratives': in general we should take the proposition expressed by ϕ in *w* to be the proposition that *would be* expressed by ϕ *if it were* used in *w*. (According to Remark 1 of Section XIX of *Demonstratives*, the 'Content of a sentence in a context is, roughly, the proposition the sentence would express if it were uttered in that context' [Kaplan 1977, 546].) One problem with this suggestion is that, on just about anyone's conception of worlds, it is a non-contingent matter which sentences are used in which worlds (and more generally, what is so in a given world), wherefore a sentence that is not used in a world could not have been used in that world. (Kaplan's contexts, I should note, are not simply worlds: they also include agents, times, and locations. Nevertheless, the same worry applies: given that it is impossible for an utterance to occur in *w* unless it does occur in *w*, it is also impossible for an utterance that doesn't occur in *w* to occur in $\langle w, a, t, l \rangle$ if this means, as it is naturally interpreted, that the utterance occurs in *w*, and is, in *w*, produced by agent *a* at time *t* and location *l*.)
10. One rarely encounters a compositional treatment of variable-binding: see Yli-Vakkuri (2013).
11. As in, e.g. Kaplan (1977).
12. The assumption that there is an intended model is something of an idealization, *inter alia*, for the simple reason that there may be too many entities being theorized about (e.g. points of evaluation) to form a set — a well-known general limitation of set-theoretic semantics, which I will set to one side.
13. See Williamson (1994, Section 5) and (1999).
14. Caie (2012, 365).
15. *Ibid*.
16. See note 23.
17. See Davies and Humberstone (1980).

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

18. Kaplan's contexts, however, are not simply worlds: see note 9.
19. The example is inspired by Williamson (1994, 253–254). I use it because it is difficult to come up with other kinds of uncontroversial examples of borderline identity statements involving proper names. Many of the standard examples are arguably definite cases where the source of the intuition of vagueness is in the vagueness of, say, 'is located at' or 'is part of.'
20. Thanks to Peter Fritz and Jeremy Goodman for discussion here.
21. Cresswell (1990) takes just this approach. There are no expressions in his quantified modal–temporal language whose interpretation is fixed in all of his models: consequently, 'there will be no wff true in all interpretations, and thus no intensional logic.' Yet he says of this language that 'among its possible interpretations is one which comes closest to reflecting the meanings of particular words in a particular natural language, for convenience English' (7).
22. Caie's (2012) main counterexample to otherworldly semantics exploits this flaw, and some of the arguments in Kearns and Magidor (2008) could be reconstructed as doing so, although the latter are not directly concerned with epistemicist interpretations of $\triangle$.
23. Caie and Hawthorne both argue that Williamson's epistemicism has problems accounting for the definiteness of certain disquotational sentences. Both arguments, however, rely on premises that seem to me less secure than the ones I use.

    Caie (2012, Section 5) does not discuss T-sentences but what I call (below) *R-sentences*: sentences of the form '"*N*" refers to *N*.' The questionable assumption in Caie's argument is that (schematically), for some name '*N*' there is a close world in which '*N*"s referent is different from its actual referent but in which 'refers' expresses the same content as it actually does. (An analogous assumption about sentences and the truth predicate could be used to argue for the non-definiteness of some healthy T-sentences.) But Hawthorne's (2006) 'domestic stability' solution seems to involve rejecting this assumption.

    Hawthorne's (2006, Section 13) argument concerns T-sentences and involves what seems to me a questionable move from the plausible claim (given Semantic Plasticity) that the truth predicate expresses a content different from its actual content in some close worlds to the further claim that there is a close world where the content the truth predicate has in that world determines an extension different from the extension determined in that world by the actual content of the truth predicate (197). There may be a good argument for why an epistemicist committed to Semantic Plasticity would have to accept the latter claim, but Hawthorne does not offer one.
24. Suppose that (a) $|\phi|_{@,w} \neq |\phi|_{w,w}$, (b) $@Rw$, and that (c) ϕ is suitable. Note that

$$\begin{aligned} |\phi|_{w,w} = 1 \quad &\text{iff} \quad \phi \in [\![T]\!](@, w) && \text{by (1) and (c)} \\ &\text{iff} \quad [\![\ulcorner \phi \urcorner]\!](@, w) \in [\![T]\!](@, w) && \text{by (ix)} \\ &\text{iff} \quad |T(\ulcorner \phi \urcorner)|_{@,w} = 1 && \text{by (viii),} \end{aligned}$$

    so $|\phi|_{w,w} = |T(\ulcorner \phi \urcorner)|_{@,w}$. Then, by (a), $|\phi|_{@,w} \neq |T(\ulcorner \phi \urcorner)|_{@,w}$, so $|T(\ulcorner \phi \urcorner) \equiv \phi|_{@,w} = 0$. By (b) and either ($A_1$) or ($A_2$), $|\triangle(T(\ulcorner \phi \urcorner) \equiv \phi)|_{w,w} = 0$. Note that this argument requires the assumption that *R* is symmetric. There is, I think, an equally plausible argument for the existence of healthy T-sentences that fail to be definite in some close worlds that does not require the assumption that *R* is symmetric: we can simply assume that, in the case at hand, both $wR@$ and $@Rw$. I take it to be fairly clear that there will be such cases if one or another of the actualistic semantics is correct, but I won't supply the philosophical argument for this for lack of space.

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

25. $|Ref(\ulcorner\tau\urcorner, \tau)|_{@,w} = 1$ iff $\langle [\![\ulcorner\tau\urcorner]\!](@, w), [\![\tau]\!](@, w)\rangle \in [\![Ref]\!](@, w)$ by (viii)
iff $\langle \tau, [\![\tau]\!](@, w)\rangle \in [\![Ref]\!](@, w)$ by (ix)
iff $[\![\tau]\!](w, w) = [\![\tau]\!](@, w)$ by (3)

so
$|Ref(\ulcorner\tau\urcorner, \tau)|_{@,w} = 0$ iff $[\![\tau]\!](w, w) \neq [\![\tau]\!](@, w)$,
so, in particular,
(3*) If $[\![\tau]\!](w, w) \neq [\![\tau]\!](@, w)$ and $wR@$ then $|\triangle Ref(\ulcorner\tau\urcorner, \tau)|_{w,w} = 0$.

26. It is not entirely clear to me that Caie's two options are the most natural ones. In fact, his taxonomy of truth conditions for $\triangle\phi$ at proper points omits the one that seems to me to most closely follow the letter of Williamson's account of ignorance of borderline matters. Consider, for example:

> What distinguishes vagueness as a source of inexactness is that the margin for error principles to which it gives rise advert to small differences in meaning, *not to small differences in the objects under discussion* (Williamson 1994, 231).
>
> Suppose I am on the 'thin' side of the boundary but only just. […] The sentence 'TW is thin' is true, but could easily have been false *without any change in my physical measurements or those of the relevant comparison class* (Williamson 1994, 231).

The most straightforward way of extrapolating a truth condition for $\triangle\phi$ at a proper point from remarks like these would seem to be this: $\triangle\phi$ is true at $\langle w, w\rangle$ iff, for all v such that wRv and the truth value of the proposition ϕ expresses in w is the same in w and in v, ϕ is true at $\langle v, v\rangle$. This is consistent with two truth definition clauses:

(O'_1) $|\triangle\phi|^{\mathfrak{U}}_{w,v} = 1$ iff, for all u such that vRu and $|\phi|^{\mathfrak{U}}_{v,u} = |\phi|^{\mathfrak{U}}_{v,v}$, $|\phi|^{\mathfrak{U}}_{u,u} = 1$.

(O'_2) $|\triangle\phi|^{\mathfrak{U}}_{w,v} = 1$ iff, for all u such that wRu and $|\phi|^{\mathfrak{U}}_{w,u} = |\phi|^{\mathfrak{U}}_{w,w}$, $|\phi|^{\mathfrak{U}}_{u,u} = 1$.

(O'_1) and (O'_2) share the logical shortcomings, respectively, of (O_1) and (O_2), and these are accompanied by similar semantic shortcomings. Unlike (O_1) and (O_2), however, (O'_1) and (O'_2) both entail that a borderline sentence expresses, in some close world, a proposition different from the one it actually expresses. This is another reason why (O'_1) and (O'_2) seem to me to be closer matches, within the 2D framework, to the letter of Williamson (1994) than (O_1) or (O_2).

27. See Williamson (2003) for discussion.
28. See Williamson (1994, Section 9.3).
29. Thanks to Peter Fritz for this evocative label.
30. I am inclined to think that we should not require R to be symmetric, but the assumption that R is symmetric helps simplify some arguments in this paper.
31. In addition to note 30, there are some delicate issues in this area, such as the status of the principles $\triangle\Box\phi \supset \Box\triangle\phi$ and $\Box\triangle\phi \supset \triangle\Box\phi$, both of which the model theory validates, which I discuss in my work with Litland. A model theory in which $\triangle$ is interpreted by a relation on $W \times W \times W$ rather than on W offers a more flexible framework for logical and semantic theorizing about L_{MV}. For maximal generality we can also add to the models a second accessibility relation to interpret $\Box$. My work with Litland explores model theories with both features.
32. For this idea to have any hope of working, we must not be too permissive about what to count as a fact 'about character.' For example, it is, in some intuitive sense, a fact about the character of 'Tim is thin' that the character of 'Tim is thin' is its actual character. Furthermore, it is a known fact. But then it cannot be a fact about

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

character in the intended sense, because if it were, then no world in which 'Tim is thin' has a character different from its actual character will be *R*-related to the actual world; in any such world the known fact that the character of 'Tim is thin' is its actual character fails to obtain.

33. This principle is plausible on it own, but it also becomes a theorem under the metasemantic interpretation of the 3D model theory, if we augment the model theory by assigning metasemantic characters to the logical constants in any of the obvious ways. Then all of the constants except for *A* will have constant characters in every metasemantic context of every model. (We would also have to add functors and an identity predicate to the language to handle the Meter Sentence.)
34. See Remark 11 of Section XIX of *Demonstratives* (Kaplan 1977, 551) for a classic articulation of the consensus view.
35. In Hawthorne's original example, we introduce an indexical singular term, 'Actuality,' and a non-indexical singular term, 'Worldy,' to designate the actual world. The identity 'Worldy = Actuality' is then known but, given natural assumptions about how the 3D model theory would have to deal with singular terms and identity, is not definite according to the metasemantic interpretation.
36. See note 8.
37. It is much less clear that I am in a position to know this much about the supervenience of content facts on 'metaphysical groundfloor' facts, for any natural interpretation of that phrase. For example, it is difficult to see how I could be in a position to know anything at all about how the content facts supervene on the microphysical facts, unless the relevant notion of 'being in a position to know' packs in the idealization I have access to Chalmers''cosmoscope' or a similar device (see Chalmers 2012, 114–115).
38. Of course, models for epistemic logic do not usually come with a designated actual epistemic possibility, nor with contexts that determine epistemic possibilities (because they usually do not come with contexts at all), but this is just an artifact of the model theory. A semantics must recognize, for each parameter shifted by an operator, a function that assigns to each context the value of that parameter that is 'realized' or 'present' (etc.) in that context.
39. It would be natural to strengthen this condition by adding the conjunct: 'and *w* matches actuality with respect to the truth value of the proposition ϕ actually expresses,' because possibilities in which one is in error because the facts that actually ϕ concerns (e.g. whether Tim is thin) are different than they actually are irrelevant to vagueness-induced ignorance (see Williamson 1994, 231). However, even the weaker condition does not hold.
40. Specifically, we must assume that the logic of $\triangle$ validates each instance of the *K* axiom schema $\triangle(\phi \supset \psi) \supset (\triangle\phi \supset \triangle\psi)$, and that $\triangle\phi$ is valid whenever ϕ is a truth-functional tautology — call this latter principle *Tautological Definitization*. Because $\triangle\sharp$ is true and $\triangle\chi$ is not, by K, $\neg\triangle(\sharp \supset \chi)$ is true. Furthermore, because χ is borderline, $\neg\triangle\neg\chi$ is true, and we can show that $\neg\triangle\neg(\sharp \supset \chi)$ is true as follows. First we assume for a *reductio*:

    (i) $\triangle\neg(\sharp \supset \chi)$.

    (ii) is valid by K.

    (ii) $\triangle(\neg(\sharp \supset \chi) \supset \neg\chi) \supset (\triangle\neg(\sharp \supset \chi) \supset \triangle\neg\chi)$,.

    and (iii) is valid by Tautological Definitization.

    (iii) $\triangle(\neg(\sharp \supset \chi) \supset \neg\chi)$.

    Because χ is borderline,

    (iv) $\neg\triangle\neg\chi$.

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

(i), (ii), and (iii) imply $\triangle\neg\chi$, which contradicts (iv). We get:

$$\neg\triangle\neg(\sharp \supset \chi) \wedge \neg\triangle(\sharp \supset \chi),$$

which is definitionally equivalent to $\nabla(\sharp \supset \chi)$.

41. This argument does not essentially depend on the assumption that ♯ expresses all of the *use* facts; any kind of fact (e.g. microphysical) on which content facts supervene will do. Nor does the argument require the assumption that ♯ specifies *all* of the facts of the relevant kind; it is enough that ♯ specifies enough of them for it to be the case that, necessarily, if the proposition actually expressed by ♯ is true, then χ expresses $p_{@}$. Thanks to John Hawthorne for discussion here.
42. While I find this argument to be pretty decisive, I should note that at least one philosopher sympathetic to epistemicism rejects two of its assumptions. Magidor (forthcoming) rejects one of the 'very minimal assumptions about the logic of definiteness' mentioned in the main text, namely the *K* axiom for $\triangle$ (see note 40), and Kearns and Magidor (2012) reject the supervenience of content facts on use facts (as well as on microphysical facts, and apparently on any facts other than the content facts themselves), so they would reject (**).
43. Williamson himself appears to posit such ignorance. One relevant passage is this:

> Since the content of the concept depends on the overall pattern, you have no way of making your use of a concept on a particular occasion sensitive to the overall pattern. Even if you did know all the details of the pattern (which you could not), you would still be ignorant of the manner on which they determined the content of the concept (1994, 231–232).

See also Williamson (1994, Section 7.4).

44. Williamson (1994, 230).

Acknowledgments

I would like to thank Catharine Diehl, Cian Dorr, Kit Fine, Peter Fritz, Jeremy Goodman, John Hawthorne, Jon Litland, Mark McCullagh, Ofra Magidor, Beau Madison Mount, Jeff Sanford Russell, Gabriel Uzquiano, Sara Kasin Vikesdal, Tim Williamson, and audiences at the University of Oxford and at the 2015 *Williamson on Modality* workshop in Montreal for helpful discussions and for comments on earlier drafts of this paper.

Notes on contributor

Juhani Yli-Vakkuri is a Postdoctoral Research Fellow in Philosophy at the Centre for the Study of Mind in Nature (CSMN) at the University of Oslo. His research interests include the philosophy of language, the philosophy of mind, philosophical logic, epistemology, and metaphysics. His most recent publication is *Narrow Content*, a monograph in the philosophy of mind co-authored with John Hawthorne and forthcoming from Oxford University Press in 2017.

References

Caie, M. 2012. "Vagueness and Semantic Indiscriminability." *Philosophical Studies* 160: 365–377.

Chalmers, D. 2012. *Constructing the World*. Oxford: Oxford University Press.

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press

Cresswell, M. 1990. *Entities and Indices*. Dordrecht: Kluwer.
Davies, M., and L. Humberstone. 1980. "Two Notions of Necessity." *Philosophical Studies* 38: 1–30.
Fritz, P. 2016. "Appendix to Juhani Yli-Vakkuri's 'Epistemicism and Modality'." *Canadian Journal of Philosophy*, this volume.
Hawthorne, J. 2006. "Epistemicism and Semantic Plasticity." In *Metaphysical Essays*, 185–210. Oxford: Oxford University Press.
Kaplan, D. 1977. "Demonstratives." Mimeograph, Department of Philosophy, UCLA. Published in J. Almog et al., eds., *Themes from Kaplan*. Oxford: Oxford University Press, 1989. Page references to the latter.
Kripke, S. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.
Kearns, S., and O. Magidor. 2008. "Epistemicism About Vagueness and Meta-linguistic Safety." *Philosophical Perspectives* 22: 277–304.
Kearns, S., and O. Magidor. 2012. "Semantic Sovereignty." *Philosophy and Phenomenological Research* 85: 322–350.
Litland, J. and J. Yli-Vakkuri 2016. "Vagueness and Modality." *Philosophical Perspectives*.
Magidor, O. Forthcoming. "Epistemicism, Distribution, and the Argument from Vagueness." *Noûs*.
Shapiro, S. 2007. *Vagueness in Context*. Oxford: Oxford University Press.
Sorensen, R. 1988. *Blindspots*. Oxford: Oxford University Press.
Williamson, T. 1994. *Vagueness*. London: Routledge.
Williamson, T. 1999. "On the Structure of Higher-order Vagueness." *Mind* 108: 127–143.
Williamson, T. 2003. "Vagueness in Reality." In *The Oxford Handbook of Metaphysics*, edited by M. Loux and D. Zimmerman, 690–716. Oxford: Oxford University Press.
Yli-Vakkuri, J. 2013. "Propositions and Compositionality." *Philosophical Perspectives* 27: 526–563.

https://doi.org/10.1080/00455091.2016.1201878 Published online by Cambridge University Press