Edgington on Possible Knowledge of Unknown Truth[1]

(to appear in John Hawthorne and Lee Walters, eds., *Conditionals, Probability, and Paradox: Themes from the Philosophy of Dorothy Edgington*, Oxford: Oxford University Press)

Timothy Williamson

Abstract: The paper is a response to Dorothy Edgington's article 'Possible knowledge of unknown truth' (*Synthese*, 2010), where she defends her diagnosis of the Church-Fitch refutation of the principle that all truths are knowable and analogous refutations of analogous principles, in response to my earlier criticisms of her diagnosis. Using counterfactual conditionals, she reformulates the knowability principle and its analogues to withstand Church-Fitch objection. In the present paper, I argue that in order to avoid a kind of trivialization, Edgington needs to supply a more general constraint on how the knower is allowed to specify a counterfactual situation for the purposes of her reformulated principles, and that it is unclear how to do so. I also question the philosophical motivation for her reformulation strategy, with special reference to her application of it to Putnam's epistemic account of truth. In passing, I question how dangerous Church-Fitch arguments are for analogues of the knowability principle with non-factive evidential attitudes in place of knowledge. Finally, I raise a doubt about the compatibility of Edgington's reformulation strategy with her view that counterfactual conditionals lack truth-conditions.

1. Many philosophers have been tempted by something like the idea that all truths are knowable. The idea is naturally formalized thus:

Knowability $\forall P\ (P \rightarrow \Diamond KP)$

Here $\Diamond$ and K stand for 'it is metaphysically possible that' and 'at some time someone knows that' respectively; → is just the material conditional, and the variable P takes sentence position.

Alonzo Church discovered the following refutation of Knowability. The principle entails its special case where the conjunction $P\ \&\ \neg KP$ is substituted for P. The consequent of the resulting instance is $\Diamond K(P\ \&\ \neg KP)$. But $K(P\ \&\ \neg KP)$ is impossible, because it entails both KP (since knowledge of a conjunction entails knowledge of its conjuncts) and $\neg KP$ (since knowledge entails truth). That refutes the consequent. Hence the special case reduces to $\neg(P\ \&\ \neg KP)$, the negation of its antecedent, and so to $P \rightarrow KP$. Thus Knowability entails the apparently stronger claim that every truth *is* known by someone at some time.[2] But that claim is silly. No one will ever know how many spiders were in my room exactly ten years ago. Therefore Knowability is false. As an anonymous referee for *The Journal of Symbolic Logic*, Church communicated his proof to Frederic Fitch, who was the first to publish it (Church 1945, Fitch 1963, Salerno 2009).

In 1985, Dorothy Edgington published a seminal reconsideration of the Church-Fitch argument (before its connection with Church emerged). While conceding that the argument refutes Knowability as formalized above, she explained a variant reading of the claim 'Every truth is knowable' that the argument does not refute. It depends on a distinction between the situation in which one knows and the situation one knows about. More specifically, her reading is this:

E-Knowability $\forall P\ \forall s\, ((\text{in } s\text{: } P) \rightarrow \exists s^*\, (\text{in } s^*\text{: } K(\text{in } s\text{: } P)))$

Here the variables *s* and *s** range over possible situations, possibilities that need not be maximally specific. E-Knowability says that whatever holds in a possible situation can be known to hold in that situation, but the knowing itself may take place in another possible situation. If, as before, one substitutes *P & ¬KP* for *P*, but in E-Knowability rather than Knowability, the consequent of the result says that in some possible situation *s** it is known that, in the possible situation *s*, *P* is an unknown truth. That involves no contradiction. Edgington's 1985 paper is clearly one of the most original and thought-provoking treatments of the Church-Fitch arguments, and deservedly one of the most cited.

A couple of years later, I published a response to Edgington's paper (Williamson 1987b). My central objection was that since E-Knowability quantifies into an epistemic context — the variable *s* occurs free in the scope of the operator *K* — we need some suitable constraint on how the knower in *s** is to specify the possible situation *s*, but none has been supplied. For one *can* quite naturally specify a situation by stating what is true in it ('the possibility that my ticket wins the lottery'), but if E-Knowability allows such specifications it becomes in a way trivial, since its consequent is verified by a possible situation in which someone knows the triviality that in the situation in which various things including that *P* are the case, it is the case that *P*.

Edgington also gave a variant of E-Knowability in notation more like that of Knowability, using modal operators including ‘actually’ (@), rigidly pointing back to the actual world, rather than quantification over non-specific situations:

EA-Knowability $\forall P\ (@P \rightarrow \Diamond K@P)$

EA-Knowability says that whatever holds in the actual world can be known in some world or other to hold in the actual world. It finesses the Church-Fitch objection to Knowability just as E-Knowability does, but is open to a similar objection. If a knower in a counterfactual world knows that in the actual world it is the case that *P*, how is she to specify the actual world? If she says ‘the actual world’ she specifies *her* world; but she was supposed to specify *our* world. If EA-Knowability allows the knower to specify the actual world by what is true in it, its consequent is verified by a possible world in which someone knows the triviality that in the world in which various things including that *P* are the case, it is the case that *P*, although in practice it may be impossible to specify all those various things in detail. Since Edgington treats human cognition as more concerned with unspecific possibilities than with maximally specific possible worlds, she concentrates on E-Knowability rather than EA-Knowability. I will do the same.

Edgington’s 1985 paper is rich in examples. Her knowers specify an alternative possibility non-trivially by means of a counterfactual conditional: the possibility that would have obtained if such-and-such had been the case. The trouble is that providing a non-trivial verification of a claim does not show that the claim does not also have a trivial verification. Rather, one needs to clarify the claim so as to exclude trivial verifications. In the case of E-Knowability, that requires something like a general constraint on which ways of specifying an alternative situation are to count. Part of my critique involved showing that features of Edgington's examples that might seem promising candidates for extrapolation to a general constraint do not in fact generalize as required (Williamson 1987a; 1987b; 2000, pp. 290-301).

In 2010 Edgington published a long-awaited reply to my critique. Part of her reply is that her general approach does not commit her to those failed extrapolations, because she can treat my examples in other ways. I quite agree. But the point of my examples was to bring out ways in which her examples do less than one might think to adumbrate a general constraint of the sort required. Just by rejecting various incorrect generalizations of her cases, Edgington does not thereby provide a correct generalization of them.

Section 2 will discuss in more detail the problem of advancing from Edgington's discussion of her examples to a suitably general clarification of E-Knowability. In particular, it explains why just stipulating that the knowledge at issue must take some non-trivial form, without further specifying that form, does not solve the problem. Of course, the clarification of E-Knowability should fit the intended philosophical point of proposing the principle. Section 3 questions Edgington's underlying motivation for her strategy of reformulating Knowability and similar principles, with special reference to her application of it to an epistemic account of truth once defended by Hilary Putnam. Section 4 briefly queries the relation between Edgington's general view of counterfactual conditionals and her use of them in defence of her reformulated principles.

2. Edgington makes clear that when she treats knowledge of a counterfactual as constituting knowledge *de re* about a possibility, the possibility she means is typically not just the possibility that so-and-so literally expressed by the antecedent, but rather a more specific possibility that would have obtained if so-and-so. For example, she says ‘to have enough handle on *which* possibility one is talking about, one refers to it as the one that would have developed, had there been a course of history which diverged at a certain point from the actual history’ (Edgington 2010, p. 48). The courses of history she needs include something’s being an unknown truth, but she does not write that into the antecedent of the counterfactual conditional itself, on pain of trivializing the knowledge at issue in E-Knowability.[3]

Some notation will facilitate the discussion. Edgington’s possibilities are possible *situations*. We also allow impossible situations (although the quantifiers in E-Knowability do not range over them). A situation s is possible if and only if it is metaphysically possible for s to obtain. One way of specifying situations is by nominalizing sentences: the situation that P obtains if and only if P. A situation s *strictly implies* a situation s' if and only if it is metaphysically necessary that if s obtains then s' obtains. A situation s' is *less specific* than a situation s if and only if s strictly implies s' but s' does not strictly imply s. A possible *world* is a maximally specific possible situation, that is, a possible situation s no less specific than any possible situation. The locution ‘in s: P’ is equivalent to ‘the situation s strictly implies the situation that P’. For present purposes, we may take ‘situation’ to apply to coarsely individuated items: situations are identical if and only if they are mutually strictly implying. In other words, $s = s'$ if and only if it is metaphysically necessary that s obtains if and only if s' obtains.

We also assume that any situations have a conjunction, which (as a matter of metaphysical necessity) obtains if and only if all of them do. A situation *s* *counterfactually implies* a situation *s′* if and only if had *s* obtained, *s′* would have obtained.[4] Since such counterfactual conditionals may be contingent, which world we evaluate them with respect to makes a difference. For convenience, we take the default world of evaluation as a fixed possible world in which the putative knowing takes place. ‘Counterfactually implies’ should be understood accordingly.

One obvious objection to specifying a possibility as the one that would have obtained if so-and-so is that the definite description is improper. Many different situations would have obtained if so-and-so. In the notation just introduced, the situation that so-and-so counterfactually implies many different situations. For if a situation *s* counterfactually implies a situation *s′*, then *s* also counterfactually implies any situation less specific than *s′*. In particular, any situation counterfactually implies both itself and the trivial situation that necessarily obtains. Thus any non-trivial situation counterfactually implies at least two situations. Moreover, on most views of counterfactuals even the trivial situation counterfactually implies at least two situations, since it counterfactually implies both itself and a slightly non-trivial situation that obtains in all but some very remote possibilities.

The obvious fix is to specify a situation as ‘the *most specific* situation counterfactually implied by *s*’, in other words, the situation counterfactually implied by *s* that strictly implies all situations counterfactually implied by *s*. Clearly, if s_1 and s_2 are both situations counterfactually implied by *s* that strictly imply all situations counterfactually implied by *s*, then s_1 strictly implies s_2 and *vice versa*, so $s_1 = s_2$ by the coarse-grained criterion of identity for situations above. Thus uniqueness is assured.

However, as so often, securing uniqueness jeopardizes existence. Why must there be a most specific situation counterfactually implied by the given situation *s*? There is a most specific situation counterfactually implied by any given situation if and only if the counterfactual conditional commutes with conjunction, in the sense that a situation counterfactually implies some situations if and only if it counterfactually implies their conjunction. For suppose that the counterfactual conditional commutes with conjunction. Then *s* counterfactually implies the conjunction of all those situations it counterfactually implies separately. That conjunction is the most specific situation counterfactually implied by *s*. Conversely, suppose that there is a most specific situation counterfactually implied by any given situation. Let *s* be a situation, and s^+ the most specific situation counterfactually implied by *s*. Consider some situations *S*. Suppose that *s* counterfactually implies each one of *S*. Then s^+ strictly implies each one of *S*, and so strictly implies the conjunction of *S* (*strict* implication evidently commutes with conjunction), so *s* counterfactually implies the conjunction of *S*. Conversely, suppose that *s* counterfactually implies the conjunction of *S*. Then s^+ strictly implies the conjunction of *S*, and so strictly implies each one of *S*, so *s* counterfactually implies each one of *S*. Therefore the counterfactual conditional commutes with conjunction.[5]

Although the commutativity of the counterfactual conditional with conjunction looks obvious, it is invalid in some mainstream logics of counterfactuals. In particular, David Lewis's preferred semantics for the counterfactual conditional makes it commute with finite conjunctions but not with infinite ones (Lewis 1973, pp. 19-21 and 132). In his terminology, the Limit Assumption may fail. Notoriously, Lewis allows cases such as this: for every positive length *l*, if this line had been longer then it would have been longer by less than *l* (for it would have been longer by at most *l*/2); but it is of course false that if this line had been longer then, for every positive length *l,* it would have been longer by less than *l* (for that is tantamount to saying that if it had been longer it would not have been longer).[6] Thus for Lewis, even though this line could have been longer, there is no such thing as the *most specific* situation that would have obtained if it had been longer.

On a crude probabilistic account, a counterfactual conditional is true if and only if the chance of the consequent conditional on the antecedent exceeds a fixed threshold *c* less than 1. Then the counterfactual conditional fails to commute even with finite conjunctions. If the probabilities are doing much work, such an account may make it quite rare for there to be a most specific situation counterfactually implied by a given situation.

On Stalnaker's theory of conditionals (1968), every possibility counterfactually implies a unique possible world, which is the most specific situation (and a maximally specific possible situation) counterfactually implied by that possibility. This seems at odds with Edgington's emphasis on unspecific possibilities, although the contrast may be blurred by Stalnaker's qualification that it is often indeterminate *which* possible world a given possibility counterfactually implies (Stalnaker 1984, pp. 132-46).

Even if most possibilities do not counterfactually imply maximally specific possibilities, there may still be a most specific situation counterfactually implied by any given situation, for the counterfactual conditional may still commute with conjunction, as in the Lewis semantics with the Limit Assumption imposed. Commutativity will typically hold if the counterfactual conditional is a strict conditional restricted to contextually relevant worlds.[7]

That there always is a most specific situation counterfactually implied by a given situation is thus a contested assumption. Nevertheless, I will grant it to Edgington, because the commutativity of the counterfactual conditional with conjunction is a very plausible and attractive principle.[8] Those sceptical of it may still grant it to Edgington for the sake of argument.

Let us return to the Edgingtonian proposal that knowledge of the counterfactual conditional that if *A*, *C* constitutes *de re* knowledge, of the most specific situation that would have obtained if *A*, that in it: *C*. Let s_A and s_C be the situations that *A* and that *C* respectively, and s_A^+ the most specific situation counterfactually implied by s_A (we have just granted that there is such a situation as s_A^+). In present notation, the proposal is that knowledge that s_A counterfactually implies s_C constitutes knowledge *de re*, of s_A^+, that it strictly implies s_C.

The two contents of putative knowledge are at least guaranteed to have the same truth-value in the knower's world. By definition of s_A^+, s_A counterfactually implies s_C in that world if and only if s_A^+ strictly implies s_C. It does not follow that the two contents strictly imply each other. For it may well be contingent whether s_A counterfactually implies s_C but non-contingent whether s_A^+ strictly implies s_C.[9] The point is that the definite description used to fix the reference of the rigid situation term 's_A^+' in the knower's world can pick out a different situation when used in another world. One may wonder how knowledge of the truth that s_A counterfactually implies s_C can constitute knowledge of the distinct truth *de re*, of s_A^+, that it strictly implies s_C. But let us grant Edgington that, somehow, it can.

Once again, a kind of triviality threatens. For suppose that someone knows that s_A counterfactually implies s_C. Thus s_A does indeed counterfactually imply s_C. Let $s_{A\&C}$ be the situation that A & C, and $s_{A\&C}^+$ the most specific situation counterfactually implied by $s_{A\&C}$. Then $s_{A\&C}^+ = s_A^+$. For in standard logics of the counterfactual conditional □→, such as those of Lewis and Stalnaker, one can easily derive this theorem:

(*) $$(A \mathrel{\square\!\!\rightarrow} C) \rightarrow ((A \mathrel{\square\!\!\rightarrow} P) \leftrightarrow ((A \,\&\, C) \mathrel{\square\!\!\rightarrow} P))$$

Informally: if C in the closest worlds in which A, then the closest worlds in which A are the closest worlds in which A & C.[10] Since s_A counterfactually implies s_C by hypothesis, s_A and $s_{A\&C}$ counterfactually imply exactly the same situations, by (*). Therefore, by definition of s_A^+ and $s_{A\&C}^+$, $s_{A\&C}^+ = s_A^+$. By assumption, knowledge that s_A counterfactually implies s_C constitutes knowledge *de re*, of s_A^+, that it strictly implies s_C. By parity, knowledge that $s_{A\&C}$ counterfactually implies s_C constitutes knowledge *de re*, of $s_{A\&C}^+$, that it strictly implies s_C. Since $s_{A\&C}^+ = s_A^+$, knowledge *de re,* of $s_{A\&C}^+$, that it strictly implies s_C *is* knowledge *de re*, of s_A^+, that it strictly implies s_C (the use of Leibniz's law here is unproblematic because the context is *de re*). Therefore knowledge that $s_{A\&C}$ counterfactually implies s_C constitutes knowledge *de re*, of s_A^+, that it strictly implies s_C. In effect, the very *de re* knowledge required to verify an instance of E-Knowability is constituted not only by the non-trivial knowledge that A □→ C, just as Edgington had in mind, but equally and independently by the utterly trivial knowledge that $(A \mathbin{\&} C) \mathrel{\Box\!\!\rightarrow} C$, a logical truth. Thus Edgington needs to gloss E-Knowability with some constraint on how the knowledge in question is constituted, in order to prevent this kind of trivialization of the principle. The constraint had better be reasonably general, in order not to invite the charge of *ad hoc* manoeuvring.

Edgington sometimes writes as though the problem were to distinguish between ways of specifying possibilities that achieve identifying reference and ways that fail to do so, where ‘identifying reference’ involves ‘knowing which possibility one refers to’ (Edgington 2010, p. 49). But ‘the most specific situation that would have obtained if $A \mathbin{\&} C$’ is in no way obviously worse than ‘the most specific situation that would have obtained if A’ at letting one know which situation one is referring to. Indeed, the former description gives one *more* explicit information than the latter about what holds in the situation. The triviality is not intrinsic to the antecedent; it lies in the relation between the antecedent and the consequent.

Could Edgington cut the Gordian knot by simply explicitly requiring the *de re* knowledge of the situation to be constituted by non-trivial knowledge of a counterfactual conditional? The danger for this suggestion is of achieving the wrong sort of non-triviality. For example, let D be a lawlike scientific hypothesis, whose truth-value it is highly non-trivial (but possible) to determine; D has no special relevance to A or C. In fact, D is deeply nomologically impossible, whereas $A \mathbin{\&} C$, though false, could very easily have been true. If the disjunction $(A \mathbin{\&} C) \vee D$ had been true, its first disjunct would have been true.[11] Thus the most specific situation that would have obtained if $(A \mathbin{\&} C) \vee D$ is the most specific situation that would have obtained if $A \mathbin{\&} C$. But for all one knows prior to scientific inquiry, D is true but $A \mathbin{\&} C$ still false, in which case the most specific situation that would have obtained if (A & C) $\vee$ D is the most specific situation that would have obtained if D (on Lewis's view, the knower's own world). Thus the very *de re* knowledge required to verify an instance of E-Knowability is constituted not only by the utterly trivial knowledge that (A & C) $\Box\!\!\rightarrow C$ but also by the highly non-trivial knowledge that ((A & C) $\vee$ D) $\Box\!\!\rightarrow C$. The trouble is that the non-triviality of the latter knowledge concerns only the falsification of D; it has nothing to do with the relation between A and C. To allow such knowledge to verify the relevant instance of E-Knowability would entirely pervert the intended philosophical significance of the principle, as capturing the spirit of the claim 'All truths are knowable'. For the target truth in this instance was C: we were to know *de re*, of a certain situation, that in it: C. But the core non-trivial knowledge we wound up with was of the falsity of D, which has no bearing on C. Knowledge of ((A & C) $\vee$ D) $\Box\!\!\rightarrow C$ is trivial with respect to knowledge of C because it derives just from the outlandishness of D.

A similar point applies to the observation that E-Knowability and EA-Knowability are not entirely trivial because they imply the possibility of entertaining any truth of the form 'in s: P'. A realist may indeed allow the possibility of truths that cannot be so much as thought, let alone known. But a principle that rules out *that* possibility does not thereby capture the spirit of the claim 'All truths are knowable'.

One moral of the discussion so far is this. Without an underlying philosophical purpose, it is unwise to become involved in the project of fine-tuning something like Knowability in the hope of finding a principle in the vicinity that is neither trivially true nor trivially false, for the project lacks direction in the absence of a standard by which to judge whether a candidate principle meets the *point* of the original claim. Let us therefore examine Edgington’s philosophical motivation.

3. In *Knowledge and its Limits*, I complained that E-Knowability and EA-Knowability do not fit the arguments given by anti-realist philosophers such as Dummett for the claim that all truths are knowable, in which they reach verificationist conclusions by analysis of alleged conditions of understanding (Williamson 2000, p. 300). Edgington concedes this point: ‘Williamson may well be right that the sort of knowability I defend will be of little comfort to those who seek a systematic, verificationist theory of meaning’. She explains: ‘I was not trying to defend knowability with that aim in mind. The holistic nature of evidential support — its strong dependence on background beliefs — makes such a project unfeasible, in my view’ (Edgington 2010, p. 51). I agree with her holistic objection to verificationist theories of meaning and understanding (it is not the only objection). But then why seek to refine Knowability?

Edgington states her general positive motivations for E-Knowability and EA-Knowability thus:

> Rather, it struck me as implausible that hosts of very mundane facts should be in principle unknowable. Also, there are certain philosophical positions which, it seemed to me, would be defeated too readily by a Fitch-like argument, and which may be consistently restated by the technique I propose. (Edgington 2010, p. 51)

I take the first reason first, the idea that it is ‘implausible that hosts of very mundane facts should be in principle unknowable’. With no verificationism or anti-realism in the background, why should we *expect* even very mundane facts to be in principle knowable? The universe was not designed to facilitate our knowing. Perhaps what makes a fact ‘very mundane’ is that it is about objects, properties, and relations that are very familiar to us, which implies that we already have easy epistemic access to them. But it does not imply that we have easy epistemic access to every combination of those objects, properties, and relations. Fitch-Church arguments concern facts that can be stated in very ordinary terms, but those terms are assembled into a subtle logical structure, involving universal quantification (over knowing subjects), negation, and an epistemic operator. Do we have any more reason to expect such a sentence not to state an unknowable fact than we have to expect the very mundane sentence ‘I spent my summer holiday in a village where one villager is a barber who shaves just those villagers who do not shave themselves’ not to state a contradiction? Roy Sorensen (1988) plausibly assimilates Church-Fitch unknowability to blindspots, themselves rather mundane phenomena. Elsewhere I have argued that ordinary limits on our powers of perceptual and reflective discrimination surround us with clouds of unavoidable ignorance of very mundane fact, for reasons quite different from Church-Fitch arguments (Williamson 1994a, 2000). For scientific purposes, it is pragmatically better for us not to give up too easily on the attempt to know, but we cannot require the universe to acknowledge our right to know even very mundane facts. Thus the weight of Edgington’s motivation needs to fall mainly on her second reason.

As an example of a philosophical position which would be defeated too readily by a Church-Fitch argument, and which may be consistently restated by her technique, Edgington offers an epistemic account of truth once defended by Putnam. I will examine this application of her technique in detail, as a test case.

Here is Putnam's thesis, in Edgington's words: 'truth cannot transcend what could be predicted by a "theory" which is ideal by pragmatic standards'. She formulates a Church-Fitch argument against Putnam's claim thus:

> Putnam would concede that we may never obtain such an ideal theory. Suppose we do not. So there may be truths which are not predicted by any theory we ever devise. Let *p* be such a truth. So '*p* and no theory ever devised predicts that *p*' is a truth. So by Putnam's thesis, there is a possible ideal theory which predicts that: *p* and no theory ever devised predicts that *p*. But how *could* an ideal theory predict that *p* and no theory ever devised predicts that *p*? On the reading generated by a parallel to Fitch's argument, such a theory makes inconsistent predictions.

She then applies her technique to restate Putnam's thesis:

> But of course there is another, consistent reading: there is a possible, non actual theory which predicts that *p* and that none of the actually devised theories predicts that *p*. (Edgington 2010, p. 51)

An initial concern with Edgington's formulation of the Church-Fitch argument is that it assumes that predicting anything of the form '*p* and no theory ever devised predicts that *p*' constitutes making 'inconsistent predictions'. But it does not constitute making *logically* inconsistent predictions. For the theory 'No theory is ever devised' is consistent by normal standards of logical consistency (it has models), even though devising it falsifies it, and it predicts (because it trivially entails) 'No theory is ever devised and no theory ever devised predicts that no theory is ever devised', which is of the form at issue. Perhaps Edgington intends a more pragmatic notion of inconsistency, such as being manifestly false-if-devised. A theory which predicts that *p* and no theory ever devised predicts that *p* is pragmatically inconsistent in that sense, and so clearly not ideal by pragmatic standards. Thus we may agree with Edgington that the Church-Fitch argument refutes Putnam's thesis on the reading the argument assumes.

To formulate the analogue of E-Knowability for Putnam's thesis, we make the appropriate substitution for *K* in E-Knowability. Here is the result:

E-Predictability $\forall P\ \forall s\ ((\text{in } s\text{: } P) \rightarrow$

$$\exists s^*\ (\text{in } s^*\text{: } \exists T\ (T \text{ is ideal } \&\ T \text{ predicts that in } s\text{: } P)))$$

Here 'ideal' abbreviates 'an ideal theory by pragmatic standards'. We may follow Edgington in assuming that a theory is ideal by pragmatic standards only if it is devised. For the analogue of the Church-Fitch argument for Putnam's thesis, she uses the conjunction '*p* and no theory ever devised predicts that *p*'. If we substitute that conjunction for *P* and ‘the actual situation’ (on the required rigid reading of ‘actual’) for *s* in E-Predictability, and discharge the antecedent, we obtain this:

(!) $\exists s^*$ (in s^*: $\exists T$ (T is ideal & T predicts that in the actual situation:

$$(p \;\&\; \forall T^* (T^* \text{ is devised} \rightarrow \neg(T^* \text{ predicts that } p)))))$$

The reading Edgington suggests in the passage quoted above differs significantly from (!). She has ‘there is a possible, non actual theory which predicts that *p* and that none of the actually devised theories predicts that *p*’, in which the first conjunct of the prediction, *p*, occurs outside the scope of ‘actually’. But the relevant instance of her principle cannot legitimately be read that way, for that instance results from substituting the whole conjunction ‘*p* and no theory ever devised predicts that *p*’ for the sentential variable *P* in the scope of ‘in *s*’, so the first conjunct (*p*) is within the scope of ‘in *s*’ just as much as the second conjunct is. Furthermore, she omits the qualification ‘ideal’ on ‘theory’, but without the conjunct ‘*T* is ideal’ E-Predictability becomes much less interesting, since virtually anything is predicted by some non-ideal theory or other (for example, by an inconsistent theory). I will assume that Edgington was just writing loosely, and would accept E-Predictability as a fair version of what she intended, so that (!) is the proper consequent.

How does Edgington’s reformulation strategy fit the underlying philosophical motivation for Putnam’s thesis? His primary motivation for the thesis in the work she cites is an argument by reductio ad absurdum against the ‘metaphysical realist’ claim that a theory may be ideal by pragmatic standards yet false (Putnam 1978, pp. 125-6). For several reasons, Edgington's choice of Putnam's thesis to illustrate her strategy is unfortunate.

First, Putnam's supposed reductio ad absurdum of metaphysical realism depends on dismissing in three lines any appeal to a causal theory of reference, because it would merely raise the question how the word 'causes' manages to secure unique reference (1978, p. 126). That has become notorious as the 'just more theory' move, and is generally, and rightly, regarded as illegitimate.[12] After all, one could similarly object to *any* account of reference (even a minimalist disquotational one) that it merely raises the question how the words in the account manage to refer. The objection depends on making utterly unreasonable demands on a theory of reference. Since Putnam's underlying motivation for his thesis collapses anyway, it is unclear why we should be trying to reformulate that thesis. To illustrate the utility of her reformulation strategy, Edgington would do well to find a better-motivated philosophical thesis.

Second, having stated his argument, Putnam says that for it 'not to be just a new antinomy', 'one has to show that there is at least one intelligible position for which it does *not* arise' (1978, p. 127). For if the upshot of the argument is inconsistent with every intelligible position, then the argument is presumably fallacious. According to Putnam, however, the upshot of the argument is consistent with one intelligible position: a verificationist theory of understanding, on the model of Dummett's (ibid.). He concludes that 'the theory of understanding has to be done in a verificationist way' (1978, p. 129). But the holistic grounds on which Edgington distances her strategy from verificationist theories of meaning tell just as strongly against verificationist theories of understanding.

Thus Putnam’s development of his thesis assimilates him to the very people to whom Edgington’s strategy ‘will be of little comfort’, as we saw her already concede.

Third, Putnam’s argument involves an appeal to, in effect, the upward and downward Löwenheim-Skolem theorems for the language of the ideal theory. Those theorems are standardly formulated for first-order non-modal languages, and fail for languages of many other sorts. Since E-Predictability states the ideal theory’s predictions using the modal operator ‘in *s*’, defined above in terms of strict implication (although Edgington envisages us as being able to achieve a similar effect using counterfactual conditionals), delicate technical issues arise for the project of extending Putnam’s argument to the richer language her reformulation requires. Thus Edgington’s reformulation takes the language out of the class to which Putnam’s own argument applies.

Fourth, E-Predictability is much too weak to capture Putnam’s central claim against metaphysical realism, which is that ideal theories are not false. But nothing in E-Predictability requires the theory *T* not to be false in the situation *s** in which it is supposed to be ideal. Not even strengthening the conditional in E-Predictability to a biconditional would achieve that. Here is a toy illustration of the point. Assume that all formulas of the form ‘in *s*: *P*’ express either necessary truths or necessary falsehoods, since they are defined as strict implications. Suppose that ideal theories predict all necessary truths and no necessary falsehoods (perhaps in highly non-trivial ways), but may predict many contingent falsehoods. For example: in a sceptical scenario Bad, one is really a brain in a vat but appears to oneself to be in a non-sceptical scenario Good; the ideal theory in Bad predicts all and only those propositions that are true in Good. Then E-

Predictability holds, as does its strengthening to a biconditional, but Putnam's central claim fails, and metaphysical realism is vindicated, because there are false ideal theories. Thus E-Predictability does not capture Putnam's central philosophical point.

If Edgington is to find a good illustration of the utility of her reformulation strategy, she will have to look further than Putnam's thesis.

At the end of her paper, Edgington follows Bernard Williams in emphasizing that one can imagine a scene without imagining oneself in that scene. As Williams notes, the point tells against Berkeley's reduction of the perceivable to the perceived, or of the conceivable to the conceived (Williams 1973b). Edgington suggests that in reducing the knowable to the known, the Church-Fitch argument makes a mistake similar to Berkeley's (Edgington 2010, p. 52). Williams' point about the imagination is surely both correct and important. But it poses no threat to the use of the Church-Fitch pattern of argument against the claim that all truths are knowable, or similar claims. After all, what happens if we use that pattern of argument against the thesis that all truths are *imaginable*? Suppose that for some number *n*, there are exactly *n* stars and no one ever imagines that there are exactly *n* stars. Then, by the Church-Fitch reading of the imaginability thesis, it is possible that at some time someone imagines the conjunction that there are exactly *n* stars and no one ever imagines that there are exactly *n* stars. So what? The supposition that at some time someone imagines the conjunction that there are exactly *n* stars and no one ever imagines that there are exactly *n* stars is perfectly consistent, and possible. Someone could have imagined that conjunction. Of course, given that imagining a conjunction involves imagining each conjunct, the supposition entails that at some time someone imagines that there are exactly *n* stars. Thus the supposition also entails that the second conjunct of the imagined conjunction is false. Hence, unsurprisingly, the stronger supposition that the conjunction is both imagined and true *is* inconsistent. But what matters is that the Church-Fitch pattern of argument does *not* reduce the imaginable to the imagined. More specifically, it does not reduce the thesis that all truths are imaginable, on the reading amenable to that pattern of argument, to the thesis that all truths are imagined.[13] If applications of the Church-Fitch pattern of argument really involve some misunderstanding analogous to the one about the imagination that Williams diagnosed, one might expect the mistake to appear when one applies the pattern to the imagination itself: but none does. If anyone is in danger of committing Berkeley's fallacy, it is the philosopher who feels tempted by an epistemic account of truth.

4. I will raise one more concern about Edgington's treatment of the Church-Fitch argument. Edgington is best known for her important and innovative work on conditionals. In particular, she is the leading proponent of the view that they are not apt to be true or false, and should instead be evaluated in terms of conditional probabilities. She applies that view to subjunctive as well as indicative conditionals, in order to give a unified treatment of all conditionals (Edgington 1995: 320-1 and 2004). How does her denial that counterfactual conditionals are truth-valued fit her use of them in her revision of the thesis that all truths are knowable and similar claims?

The question poses several problems. First, in her treatment of the original Church-Fitch argument, Edgington freely invokes knowledge-that with a counterfactual conditional content, which she requires to constitute *de re* knowledge of a possibility. But knowledge is supposed to be factive, to entail truth. How can one know that if this were the case then that would be the case if it is not *true* that if this were the case then that would be the case? A more general challenge is to explain what it means to know something with a probability-condition rather than a truth-condition. Many putative features of knowledge beyond factiveness are characterized in terms of the supposed truth-conditions of the objects of knowledge: for instance, reliability, sensitivity, and safety. They all lack obvious probabilistic analogues.[14]

Edgington herself has less sympathy for Knowability than for analogous epistemic constraints on truth in terms of non-factive evidential terms such as ‘predictable’ and ‘probable’ in place of ‘knowledge’.[15] But even in those cases her reformulation strategy faces a related challenge: to explain how the role she postulates for counterfactual conditionals with respect to logically complex sentences such as E-Predictability relates to the semantics of those sentences.

In principle, the general problem arises independently of Edgington’s reformulation strategy. For example, the sentence ‘All those who would have posed a threat to the regime if they had been given the opportunity were rounded up and shot’ is presumably intelligible, but to whom does the plural noun phrase ‘those who would have posed a threat to the regime if they had been given the opportunity’ apply if the open sentence ‘*x* would have posed a threat to the regime if *x* had been given the opportunity’ lacks truth-conditions (relative to a context and an assignment of a value to the variable *x*)? This is, of course, an instance of the classic Frege-Geach problem for non-truth-conditional accounts of sentences of some kind in terms of the alleged role of their unembedded occurrences (Geach 1960, 1965). Since conditionals often embed somewhat awkwardly, the Frege-Geach problem might seem less pressing for them: one might hope to explain the cases in which they embed well as those in which some more or less *ad hoc* interpretative strategy is available. But Edgington's reformulation strategy treatment makes the Frege-Geach problem even more pressing for her, because on her view we grasp sentences of the form 'in *s*: *P*' in effect as counterfactual conditionals, yet her reformulated principles embed such sentences in both the antecedent and consequent of a material conditional and in the scope of both universal and existential quantifiers.

Suppose that sentences of the form 'in *s*: *P*' inherit probability-conditionality and non-truth-conditionality from counterfactual conditionals. Then Edgington's official reformulations embed probability-conditional but non-truth-conditional sentences in both the antecedent and consequent of a material conditional and in the scope of both universal and existential quantifiers, and Edgington needs to tell us what such embeddings mean. Presumably, she does not hope to do by constructing a systematic theory of meaning in terms of probability-conditions rather than truth-conditions, because that would be in effect to embark on the project of constructing a systematic, verificationist theory of meaning, in this case with a probabilistic form of verification. As we have already seen, on holistic grounds she regards such a project as 'unfeasible' (Edgington 2010, p. 51). But how else is one to give a systematic, compositional semantics for such embeddings? Alternatively, if we interpret them non-compositionally, we are in effect reading the reformulated principles in some more or less *ad hoc* non-literal manner, which is an unhappy fate for what was supposed to be a canonical formulation of a significant philosophical doctrine.[16]

The non-truth-conditional treatment of ‘in *s*: *P*’ is anyway dangerous for Edgington, because it is defined as above in terms of strict rather than counterfactual implication, so the alleged non-truth-conditionality is in effect being generalized from conditionals to modal operators, with a corresponding increase in its implausibility. In particular, when *s* is the (maximal) actual situation, ‘in *s*’ is tantamount to the rigidifying ‘actually’ operator @, as in EA-Knowability, so presumably @*P* is truth-conditional if and only if ‘in *s*: *P*’ is too. To deny truth-conditions to sentences of the form @*P* is particularly implausible, since we can quite easily and naturally specify truth-conditions for them.

The alternative is that sentences of the form ‘in *s*: *P*’ and @*P* have truth-conditions. In that case, Edgington’s account requires our knowledge (or non-factive evidence) that their truth-conditions obtain to be somehow constituted by *de re* knowledge (or non-factive evidence) of a truth-conditionless counterfactual conditional, whatever such knowledge (or non-factive evidence) might be. If she could explain how such constitution is to work, the semantics of the reformulated principles would no longer pose a special difficulty for her, since it would still fall within the domain of truth-conditional semantics. But *how* could knowledge (or non-factive evidence) of something truth-conditionless constitute knowledge (or non-factive evidence) that a truth-condition obtains? One danger for Edgington in attempting to answer that question is that her answer might involve inadvertently supplying a plausible candidate truth-condition (perhaps a condition on probabilities) after all for the supposedly truth-conditionless thing — the counterfactual conditional — contrary to her view that it has none. Alternatively, reasons for not taking the explanation that way might turn out to undermine the constitution claim itself. But so far we lack even an attempt at an explanation.

In sum: reconciling Edgington's response to the Church-Fitch argument with her account of the semantics of conditionals is no easy matter, if it can be done at all.

5. In this paper I have presented a number of serious difficulties for Edgington's defence of her treatment of Church-Fitch arguments. Nevertheless, I hope that my discussion makes it clear just how rich and rewarding are her two brief papers on the issue. There is surely much more to be said about all the problems I have raised here.

Notes

[1] Dorothy Edgington and I were colleagues at Oxford from 2003 to 2006, while she held the Waynflete Chair of Metaphysics. I remember the joint graduate classes we gave in that period as some of the most enjoyable and rewarding teaching in which I have ever participated. Dorothy creates a relaxed, unthreatening, friendly atmosphere of straightforward intellectual co-operation, which (therefore rather than nevertheless) encourages everyone to aim for the highest standards of accuracy and clarity. In particular, she has a knack of finding the simplest, most perspicuous, least fancy example to make a point. These qualities come through in her writing too. She is a model for a way of doing philosophy that is deeply scientific but not in the least dehumanizing. This chapter originates in a talk given to the 2011 conference in honour of Dorothy at the Institute of Philosophy in London. The material was also presented to a class in Oxford. I thank both audiences, and in particular John Hawthorne and Jeremy Goodman, for useful comments. Special thanks to Lee Walters, who provided valuable detailed written comments on a draft of this paper — and above all to Dorothy herself, for her wonderful contributions, both in person and in writing, both in teaching and in research, to philosophy.

[2] In intuitionistic logic, the argument yields only the weaker result $P \rightarrow \neg\neg KP$. This matters because some sympathizers with Knowability, such as Michael Dummett, were motivated by a form of verificationism that replaced classical with intuitionistic logic.

Nevertheless, the argument presents significant difficulties even for such verificationists. I have discussed the argument in the intuitionistic setting (Williamson 1982, 1988, 1992, 1994b); for a more recent treatment and further references see Murzi 2010. The present chapter assumes a classical setting, to which both Edgington and I are sympathetic.

3 Edgington's take on her principle EA-Knowability is therefore radically at odds with that suggested by David Chalmers (2012, p. 31), which involves counterfactually specifying the actual world *w* by the infinite conjunction of all sentences in an imaginary canonical language true at *w*.

4 Could we have treated 'in *s*: *P*' as equivalent to 'the situation *s* counterfactually implies the situation that *P*' (rather than to 'the situation *s* strictly implies the situation that *P*')? That reading is uncharitable to Edgington, because it tends to undermine the difference between E-Knowability and Knowability. As an extreme case, if '*T*' is a tautology, Lewis's theory of counterfactuals makes 'the situation that *T* counterfactually implies the situation that *P*' equivalent to '*P*' itself, by his 'centering' axioms (6) and (7); Lewis 1973, p. 132, which Edgington 2011, p. 83, accepts. Then E-Knowability entails Knowability, and thereby succumbs to the Church-Fitch argument. Even on a weaker logic of counterfactuals and without such an extremely unspecific situation, related effects threaten. By contrast, 'the situation that *T* strictly implies the situation that *P*' is equivalent to 'it is metaphysically necessary that *P*', which yields no such collapse when substituted into E-Knowability.

[5] Williamson 2007, pp. 293-304, provides a suitable background logic of counterfactual conditionals and metaphysical modality for the argument in the text. Note the implicit use of the principle that if s_1 counterfactually implies s_2 and s_2 strictly implies s_3 then s_1 counterfactually implies s_3, which entails that an impossibility counterfactually implies anything (since an impossibility counterfactually implies itself and vacuously strictly implies anything). Although the vacuous truth of counterpossibles is somewhat controversial, see Williamson 2007, pp. 171-5, for a defence. Counterpossibles are in any case not very relevant to the purposes of Edgington's paper. The background logic has the principle that no possibility counterfactually implies an impossibility, so if we start in the realm of possibilities, neither strict nor counterfactual implication ever leads us outside that realm.

[6] Perhaps there could have been lengths other than all actual lengths, if space had been differently structured, but we may assume that if the line had been longer space would still have been structured the same.

[7] Lewis recognized such a view as an option but rejected it as defeatist (1973, p. 13). I write 'typically' rather than 'always' because the counterfactual conditional may fail to commute even with finite conjunctions on a dynamic semantics that evaluates it as a contextually restricted strict conditional at a context updated in response to the presence of the counterfactual itself in ways sensitive to its consequent. See also Gillies 2007.

[8] For a detailed defence of the commutativity principle see Fine 2012, pp. 39-45. For discussion of its relation to the Barcan and Converse Barcan formulas see Williamson 2013, pp. 127-9.

[9] Strict implication is non-contingent in the modal logic S5, where everything is either necessarily necessary or necessarily not necessary.

[10] Edgington seems to accept (*). In support of the centering principle, Edgington 2011, p. 83, says that '[g]ood reasons' for it are found in Walters 2009, where one of the two arguments for centering invokes (*) as a premise. Some authors have rejected (*); for a recent exchange see Ahmed 2011 and Walters 2011.

[11] As Edgington reminds us, speakers often treat counterfactuals with disjunctive antecedents non-standardly, taking them to mean that each disjunct counterfactually implies the consequent, but she allows in a similar case that by heavy-handed wording we can enforce the intended standard compositional reading (Edgington 2010, p. 47). That compositional reading is intended here. At a cost only in complexity, one could also replace the disjunction by its De Morgan equivalent in terms of conjunction and negation.

[12] Lewis 1984 gives an excellent detailed discussion of Putnam's argument.

[13] Of course, the thesis that all truths are imaginable may well be false for quite different reasons, because not all propositions are capable of being entertained in thought,

the negation of a proposition is capable of being entertained in thought only if the original proposition is capable of being entertained in thought, and one of the two propositions is true.

[14] See Moss 2013 for a response to something like this challenge.

[15] Edgington's treatment of extensions of the original Church-Fitch argument to non-factive epistemic operators underestimates the extent of the difficulties in making the generalization (2010, p. 43). For example, let E be the operator 'it is probable that' (in an evidential sense), and consider the thesis that every truth can be probable, on the reading a standard Church-Fitch argument against that thesis assumes ($\forall P\ (P \rightarrow \Diamond EP)$). The argument requires the absurdity of $E(p\ \&\ \neg Ep)$, in the sense if not of logical inconsistency then at least of gross implausibility, which Edgington grants. But we can model $E(p\ \&\ \neg Ep)$ in epistemic logic within a framework for evidential probability as follows (see Williamson 2000, pp. 209-37 for background). For worlds, use the integers from $-n$ to $+n$, where n is odd, with a uniform prior probability distribution. The world j is epistemically accessible from the world j just in case $j-1 \leq k \leq j+1$ (this models a subject with limited powers of discrimination). The evidence at j is the set of worlds accessible from j (they are the worlds consistent with the evidence at j). The probability of A at j is the prior conditional probability of A on the evidence at j, in other words, the proportion of those worlds accessible from j where A holds. Let EA hold at j if and only if the probability of A at j is at least 2/3. In the model, that means that EA holds at j if and only if A holds at two or three worlds accessible from j. Let p hold at all and only odd

worlds. Then *Ep* holds at all and only even worlds, since the only odd world accessible from an odd world is itself, whereas two neighbouring odd worlds are accessible from an even world. Hence *p* & ¬*Ep* also holds at all and only odd worlds, so *E*(*p* & ¬*Ep)* holds at all and only even worlds, and in particular at 0. In other worlds, it is probable that: the world is odd and it is not probable that the world is odd. Of course, strictly speaking, what is needed for a Church-Fitch argument against the thesis that every truth can be *sometimes* probable is inconsistency in some sense (perhaps pragmatic) in the supposition that it can be *sometimes* probable that: *p* and it is *never* probable that *p*. But the only promise of such an inconsistency comes from the synchronic case. One may have to be content with a Church-Fitch argument against the stronger thesis that every truth can be both probable and true. Edgington herself makes that modification for an example involving eliminativism about folk psychology (2010, p. 43), but it seems to be needed much more widely. Although the unstrengthened claim 'All truths can be probable' lacks plausibility and motivation, its main problems do not stem from the Church-Fitch reading.

[16] The failure of compositionality would be of a far more radical sort than that envisaged, for example, in Higginbotham 1986, pp. 33-37, where he argues that the semantic contribution of a conditional to a constituent is sensitive to features of the sentence in which it is embedded, but nevertheless in a systematic way.

References

Ahmed, Arif 2011: ‘Walters on conjunction conditionalization’, *Proceedings of the Aristotelian Society,* 111, pp. 113-120.

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

Dowe, Phil, and Noordhof, Paul (eds.) 2004: *Cause and Chance: Causation in an Indeterministic World*. London: Routledge.

Edgington, Dorothy 1985: ‘The paradox of knowability’, *Mind,* 94, pp. 557-568.

Edgington, Dorothy 1995: ‘On conditionals’, *Mind,* 104, pp. 235-329.

Edgington, Dorothy 2004: ‘Counterfactuals and the benefit of hindsight’, in Dowe and Noordhof 2004, pp. 12-27.

Edgington, Dorothy 2010: ‘Possible knowledge of unknown truth’, *Synthese,* 173, pp. 41 -52.

Edgington, Dorothy 2011: ‘Conditionals, causation, and decision’, *Analytic Philosophy,* 52, pp. 75-87.

Fine, Kit 2012: ‘A difficulty for the possible worlds analysis of counterfactuals’, *Synthese*, 189, pp. 29-57.

Geach, Peter 1960: ‘Ascriptivism’, *The Philosophical Review,* 69, pp. 221-225.

Geach, Peter, 1965: ‘Assertion’, *The Philosophical Review,* 74, pp. 449-465.

Gillies, Anthony 2007: ‘Counterfactual scorekeeping’, *Linguistics and Philosophy,* 30, pp. 329-360.

Higginbotham, James 1986: ‘Linguistic theory and Davidson’s program in semantics’, in LePore 1986, pp. 29-48.

LePore, Ernest (ed.) 1986: *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Oxford: Blackwell.

Lewis, David 1973: *Counterfactuals*. Oxford: Blackwell.

Lewis, David 1984: 'Putnam's paradox', *The Australasian Journal of Philosophy*, 62, 221-236.

Moss, Sarah 2013: 'Epistemology formalized', *The Philosophical Review*, 122, pp. 1-43.

Murzi, Julien 2010: 'Knowability and bivalence: intuitionistic solutions to the Paradox of Knowability', *Philosophical Studies*, 149, pp. 269-281.

Putnam, Hilary 1978: *Meaning and the Moral Sciences*. London: Routledge and Kegan Paul.

Rescher, Nicholas (ed.) 1968: *Studies in Logical Theory*. Oxford: Blackwell.

Sorensen, Roy 1988: *Blindspots*. Oxford: Clarendon Press.

Stalnaker, Robert 1968: 'A theory of conditionals', in Rescher 1968, pp. 98-112.

Stalnaker, Robert 1984: *Inquiry*. Cambridge, Mass.: MIT Press.

Walters, Lee 2009: 'Morgenbesser's coin and counterfactuals with true components', *Proceedings of the Aristotelian Society*, 109, pp. 365-379.

Walters, Lee 2011: 'Reply to Ahmed', *Proceedings of the Aristotelian Society*', 111, pp. 123-133.

Williams, Bernard 1973a: *Problems of the Self*. Cambridge: Cambridge University Press.

Williams, Bernard 1973b: 'Imagination and the self', in Williams 1973a, pp. 26-45.

Williamson, Timothy 1982: 'Intuitionism disproved?', *Analysis*, 42, pp. 203-207.

Williamson, Timothy 1987a: 'On knowledge of the unknowable', *Analysis*, 47, pp. 154 -158.

Williamson, Timothy 1987b: ‘On the paradox of knowability’, *Mind*, pp. 256-261.

Williamson, Timothy 1988: ‘Knowability and constructivism’, *The Philosophical Quarterly*, 38, pp. 422-432.

Williamson, Timothy 1992: ‘On intuitionistic modal epistemic logic’, *Journal of Philosophical Logic*, 21, pp. 63-89.

Williamson, Timothy 1994a: *Vagueness*. London: Routledge.

Williamson, Timothy 1994b: ‘Never say never’, *Topoi*, 13, pp. 135-145.

Williamson, Timothy 2000: *Knowledge and its Limits*. Oxford: Oxford University Press.

Williamson, Timothy 2007: *The Philosophy of Philosophy*. Oxford: Blackwell.

Williamson, Timothy 2013: *Modal Logic as Metaphysics*. Oxford: Oxford University Press.