



Counterfactual knowability revisited

Julian J. Schlöder¹ 

Received: 16 August 2018 / Accepted: 9 January 2019
© The Author(s) 2019

Abstract

Anti-realism is plagued by Fitch's paradox: the remarkable result that if one accepts that *all truths are knowable*, minimal assumptions about the nature of knowledge entail that *every truth is known*. Dorothy Edgington suggests to address this problem by understanding *p is knowable* to be a counterfactual claim, but her proposal must contend with a forceful objection by Timothy Williamson. I revisit Edgington's basic idea and find that Williamson's objection is obviated by a refined understanding of counterfactual knowability that is grounded in possible courses of inquiry. I arrive at a precise definition of knowability that is not just a technical avoidance of paradox, but is metaphysically sound and does justice to the anti-realist idea.

Keywords Knowability · Fitch's paradox · Knowledge · Truth · Modality

1 Introduction

Anti-realists endorse the thesis that *all truths are knowable* (Dummett 1991; Tennant 1997). The celebrated result known as Fitch's paradox appears to show that the anti-realist thesis is incompatible with the nigh-unassailable claim that *there is an unknown truth*. Many solutions to the paradox have been proposed (Tennant 1997; Beall 2000; Fara 2010; Fischer 2013; Artemov and Protopopescu 2013, to name just a few), but 'there continues to be no consensus on whether and where' things go wrong for the anti-realist (Brogaard and Salerno 2013). There is equally little consensus on just what kind of modality is the *able in knowable*.

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 758540).

✉ Julian J. Schlöder
julian.schloeder@gmail.com

¹ Institute for Logic, Language and Computation, University of Amsterdam, P.O. Box 94242, 1090 GE Amsterdam, The Netherlands

In this paper, I revisit—and attempt to vindicate—the suggestion of Dorothy Edgington (1985) that *knowability* should be understood to denote counterfactual knowledge of actual truth. This proposal has several attractive features. It is a *factive* account of knowability, i.e. all knowable propositions are actually true (as they should; cf. Fuhrmann 2014),¹ and it unrestrictedly makes *all* actual truths knowable. This allows the anti-realist to *identify* truth and knowability (as she should; cf. Kvanvig 1995). Moreover, characterising knowability as counterfactual knowledge of actuality requires no more than the usual vocabulary of counterfactual semantics, whereas some other proposals must introduce new modalities or new takes on metaphysical possibility (e.g. Rückert 2004; Tennant 2009; Fara 2010; Fischer 2013; Fuhrmann 2014).²

However, Edgington's proposal must contend with a powerful objection by Timothy Williamson (1987). He argues that there cannot be any non-trivial counterfactual knowledge of actuality and I demonstrate that a certain strengthening of this objection remains forceful against Edgington's (2010) more recent response. However, I then continue to argue that if counterfactual knowledge is defined with respect to those possibilities that arise from the counterfactual pursuit of some knowledge-imparting process, the objection is met. Appealing to knowledge-imparting processes should come as an entirely natural step for the anti-realist, who generally appeals to (idealised) processes of verification (Kvanvig 1995; Tennant 1997; Fischer 2013).

I consider my task to be one of explication. Fitch's paradox shows that the anti-realist's use of *knowable* is in need of clarification, and further objections raised against the anti-realist might require further clarification. In what follows, I go through some iterations of this process to ultimately arrive at a formulation of the anti-realist's thesis that appears to resist attempts to trivialise it. Technical vindications of Edgington's proposal have been developed (Rabinowicz and Segerberg 1994; Rückert 2004), but according to Williamson (2009) these solutions fail to 'fully answer the philosophical objections' (fn. 10). He does not elaborate how exactly they fail to meet these objections. In any case, the proposal developed here is not merely a formal avoidance of paradox, but a metaphysically and philosophically appreciable explication of what is meant by *knowable* that allows one to coherently claim that *truth is what is knowable*.

¹ Some solutions to Fitch's paradox discard the desideratum that knowability is factive. Notably, Fischer (2013) and Artemov and Protopopescu (2013) claim that only *stable truths* are knowable, which they express as $\Box p \rightarrow \Diamond Kp$, where the modal operators range over stages of verification. They argue that truths that become false in the process of verification are not knowable; conversely, then, falsities that become true at some stage count as knowable, falsifying factiveness. Since, arguably, the anti-realist position is characterised by an *identification* of truth and ideal verification (Kvanvig 1995), these accounts serve as clever defences of some epistemological thesis, but it is questionable how much they can aid the anti-realist.

² Using only established concepts should not count as a decisive advantage, but it benefits the dialectic: the realist and anti-realist can have a sure-footed debate when the anti-realist's thesis is stated in clear, well-understood and independently investigated terms. An instructive case is the debate between Williamson (2000, 2009) and Tennant (2001, 2002, 2010) that is arguably stifled by their disagreement about the logic and metaphysical properties of Tennant's *epistemic possibility* modality.

2 On the meaning of *knowable*

2.1 Counterfactual knowability

As persuasively argued by Dummett (1991) and Tennant (1997), among others, there seem to be good reasons (linguistic, metaphysical, epistemological) to suppose that *truth is just what is knowable*. I call this the *anti-realist thesis*; defenders of it are known as *anti-realists*. (It is worth pointing out that Edgington, whose proposal I defend and elaborate, does not have the goal to advance a thesis about truth (Edgington 2010, p. 51), although her ideas are useful to those who do.) Anti-realists must contend with the *Fitch objection* to their thesis. It goes as follows.

We may suppose the following uncontroversial assumptions about epistemic logic. Knowledge is factive ($Kp \rightarrow p$); knowledge distributes conjunctively ($K(p \wedge q) \rightarrow (Kp \wedge Kq)$); and if something is a logical contradiction, it is impossible ($p \vdash \perp$ entails $\vdash \neg \Diamond p$). Now consider the proposition *p and p is not known*, i.e. $p \wedge \neg Kp$, and note that $K(p \wedge \neg Kp)$ entails that $Kp \wedge K\neg Kp$ by distribution. The latter's right conjunct entails $\neg Kp$ by factiveness. Hence we have $Kp \wedge \neg Kp$, i.e. a contradiction. Thus, $\vdash \neg \Diamond K(p \wedge \neg Kp)$, i.e. it is impossible to know $p \wedge \neg Kp$. Hence, many are inclined to conclude, there is an unknowable truth—namely *p and p is not known* for some true but unknown *p*—and so anti-realism fails.

The argument is valid as it stands, but yet the anti-realist should protest. Of course our knowledge can change. So if there is an unknown truth *p*, we should say that it is unknown *in our present epistemic state s*. No similar triviality result follows from the claim that at a state that is not our present state it is known that *p and p is not known in s*. However, this solution needs to clarify what is meant by *present*.

It can easily be seen that *present* cannot simply be a time index (Kvanvig 1995). For if we were to write $K^t p$ for “*p* is known at time *t*”, we could consider the sentence *p and p is never known*, formalised as $p \wedge \forall t(\neg K^t p)$. Then, analogously to the standard Fitch objection, one can show that it is impossible to know at some time that *p and p is never known*. So, if we insist that *every truth is knowable* for a tensed sense of *knowable*, we have to accept that $p \wedge \forall t(\neg K^t p)$ is no truth. Worse still, if we take *every truth is knowable* to be a thesis about a metaphysical necessity (as we should, if it is to be understood as a thesis about the nature of truth), we need to accept that it is *impossible* that $p \wedge \forall t(\neg K^t p)$ is true. Hence, it follows that *necessarily* it is the case that $p \rightarrow \exists t(K^t p)$.

However, it is certainly possible that some truths will never be known. For instance, all sentient life may cease to exist tomorrow—long before we have discovered all truths. So whatever sense can be made of *knowable* when faced with the Fitch objection, it cannot be found in differentiating what we know *now* from what we *will* know.

As a first attempt to solve this puzzle, one can appeal to non-actual times. When we say that *actually* a truth won't ever be known, we are quantifying over the actual times of the actual world. So if we want to claim that it is knowable that *p and p won't be known at any actual time t*, then the time index *t'* at which this proposition is known is not an actual time, i.e. not among the *t*.

There are some formal difficulties with this, pointed out by Kvanvig (1995). If we want to quantify over times, we run into the challenges of quantifying into modal position. But for present purposes we can sidestep that issue by using temporal modals. That is, if F is a modal ‘at some time’, then the claim p won’t ever be known is $\neg FKp$. Then, to claim that at some time it is known that actually $\neg FKp$ is to claim that $\diamond FKA(\neg FKp)$ (for A being the *in the actual world* operator). The anti-realist may claim that the outer F makes different times accessible than the inner F (since A forces the inner F to be interpreted with respect to the actual world). This is the case in a standard modal semantics, and there is no formal or metaphysical problem with this.

However, this means that the anti-realist who thusly insists on the knowability of p and p is never known is making a *counterfactual* claim: p is knowable if p is known at some possibly non-actual time. Hence *knowable* includes a metaphysical possibility operator. The anti-realist may appeal to examples like the following to give plausibility to the counterfactual move.

(Oats). Say that I had muesli for breakfast. There is some number N such that N is exactly the number of oats in my bowl. Hence, $p = \text{Julian ate } N \text{ oats this morning}$ is a truth. But I didn’t count the oats and ate all the evidence, so this won’t ever be known in this world.³

Edgington (2010) is discontent, and quite rightly so, with any notion of knowability on which such “very mundane facts” such as *Julian ate N oats this morning* “should be in principle unknowable” (p. 51). There is a good sense in which p is knowable in (Oats): one *could* have counted, hence could have known, for a metaphysical reading of *could*.

However, it would be a mistake to therefore write the anti-realist thesis as $p \leftrightarrow \diamond Kp$ for \diamond being the usual operator for metaphysical possibility. It is *metaphysically* possible that grass is purple. Hence it is also metaphysically possible that someone *knows* that $p = \text{grass is purple}$. Hence there is a world in which Kp , hence it is the case that $\diamond Kp$, but certainly *grass is purple* is not true in the actual world, hence it should not count as knowable (Tennant 2009; Fara 2010; Fuhrmann 2014).

Dorothy Edgington (1985) therefore puts the anti-realist thesis as *every actual truth is knowable to be an actual truth*. Writing A for ‘in the actual world’, K for ‘it is known at some time’ (incorporating tense) and \diamond for metaphysical possibility, she can state the formal form of the anti-realist thesis as (AK).

$$\text{(AK)} \quad Ap \rightarrow \diamond KAp.$$

This avoids the Fitch objection because KA is not a factive operator in non-actual worlds. If it is actually true that p and p won’t be known, (AK) entails that $\diamond KAp$ (and p won’t be known), which is consistent because in non-actual worlds, the complex operator KA is not factive. Hence KAp and $KA(p$ won’t be known) are consistent in non-actual worlds. Moreover, $\diamond AK$ is a factive notion of knowability.

³ This example is broadly similar to some examples used by Fara (2010). He uses such cases to defend that *knowable* derives from having certain *capacities*: I have the capacity to count, hence the capacity to know, even if I in fact did not count and hence won’t ever know.

In worlds where, say, grass is purple, it cannot be known that *actually* grass is purple, since grass is actually green.

But is (AK) philosophically sound? According to the (AK) conception of knowability, a proposition p being knowable means that in some *non-actual* world someone has knowledge *about the actual world*. To seriously defend this, the anti-realist must seemingly endorse a strong principle of transworld awareness. What is more, she appears to stipulate that knowers in non-actual worlds are able to determine which world is actual. But if they are, it seems their knowledge of the actual world would be trivial. This can be seen as follows (Rabinowicz and Segerberg 1994; Artemov and Protopopescu 2013).

Generally, it holds for the actuality operator A and any proposition p that $\Box Ap \leftrightarrow Ap$. Then, a derivation of triviality from (AK) follows model-theoretically, assuming standard semantics for K (i.e. that Kp is true if at all epistemically accessible worlds, p is true). If p is any actual truth, then Ap is necessary. That is, if p is an actual truth, Ap is true at all worlds v —including at those that are epistemically accessible from the actual world. Hence, KAp is actually true. Thus $Ap \rightarrow KAp$ for any p . Call this the *simple triviality argument*.

Clearly, the anti-realist must *further* clarify her thesis to make it plausible that inhabitants of non-actual worlds can have non-trivial knowledge of actuality. Edgington (1985) already makes the useful suggestion that the operator A is not centred on the actual world, but on the actual *situation*. The actual situation consists of the actual facts relevant to the current inquiry, but does not fully specify the actual world (e.g. if the relevant proposition is about which face a coin shows, there are only two relevant situations: *heads* and *tails*).

Then, non-actual knowers can have knowledge of actuality merely by being acquainted with a specifying description of the actual situation. One may therefore explicate (AK) to the statement (αK).

$$(\alpha K) \quad Ap \rightarrow \Diamond K(\alpha \rightarrow p)$$

where α specifies the actual *situation*.

The plausibility of (αK) stands and falls with how much information is contained in α . Obviously, we cannot call a p knowable whenever there is *some* specification α of the actual situation for which it holds that $\Diamond K(\alpha \rightarrow p)$. This would be a trivial notion of knowability, since we could just include the fact that p in α . If p is part of α , then $K(\alpha \rightarrow p)$ is knowledge of a triviality, not of actuality (Williamson 1987; Fara 2010). By the very same argument, the anti-realist must reject that α logically entails p . Possible knowledge of a logical fact also does not entail anything about the actual world.

Hence, the anti-realist must answer a tough question: what is it for an α to specify the actual situation in the sense of (αK)? More specifically, if α does not entail p , how can someone in a non-actual world know that in the actual situation (that she only knows by being acquainted with α) it is the case that p ?

In the coin flip case, if the actual situation is *tails*, the relevant α must specify some facts that allow one to distinguish *tails* from *heads*, without entailing *tails* outright. The question is, what are these facts? To untangle this, it helps to consider an example due to Edgington (1985).

(Comet). Suppose there is a scientific experiment that can be performed exactly once, and failure to perform it results in that some knowledge remains unknown forever. To give a specific case, suppose there is a comet that is disintegrating, and we have precisely one shot at landing a probe on it to analyse its composition before it is gone. There is a chance of critical failure (say, the probe breaks on landing), which results in us never being able to analyse the comet (there is no time to send another probe before it disintegrates).

Then, let p denote the proposition describing the comets composition, s_K be the situation that is specified by *the experiment goes as planned* and $s_{\neg K}$ be the situation specified by *the experiment fails*.

Suppose moreover that $s_{\neg K}$ specifies the actual situation, i.e. that we actually don't know p and won't ever know p . Then $A(p \wedge \neg Kp)$ and $\diamond K(s_{\neg K} \rightarrow (p \wedge \neg Kp))$. The latter is because it is known in the situation described by s_K that *if the situation were $s_{\neg K}$, it would (still) be the case that p , but p would never be known*.

So, in the case of (Comet) the appropriate specification of actuality is $\alpha = s_{\neg K} = \textit{the experiment fails}$. Note that the knowledge of actuality in (Comet) is properly paraphrased as a *counterfactual* conditional evaluated in a non-actual *situation*. Thus, we should explicate (αK) further to $(\alpha K')$, using $\Box \rightarrow$ to express a counterfactual conditional.

($\alpha K'$) $Ap \rightarrow \diamond K(\alpha \Box \rightarrow p)$
 where α specifies the actual situation,
 and \diamond accesses possible situations.

That is, p is knowable if there is a possible situation s where it is known that *if things were as in the specification α of the actual situation, then p* . With respect to s , this conditional is counterfactual. We have seen in (Comet) an example for an appropriate α . But to defend $(\alpha K')$, the anti-realist should give general criteria to select an appropriate α .

Williamson (1987) argues that this cannot be done, since however one specifies α , if *in the situation specified by α , p is true*, then it is necessary. Hence, knowing that in the situation specified by α , p holds “requires no more than knowledge of a trivial logical truth” (p. 259). For the argument, suppose a standard account of counterfactual conditionals adapted to situations, i.e. $\varphi \Box \rightarrow \psi$ is true at a situation s if at the situation s' that contains φ and is closest to s with that property, ψ holds.⁴

Now, the defender of $(\alpha K')$ claims that in a situation s , knowledge that $\alpha \Box \rightarrow p$ constitutes knowledge of the actual situation. For this to be true, α needs to be specified such that for any proposition r in the actual situation, the α -situation closest to s must also contain r , since otherwise α would fail to specify the actual situation. *Ex hypothesi*, the actual situation contains p . Thus the α -situation closest to s is also the $(\alpha \wedge p)$ -situation closest to s . This goes for any s and so $\alpha \Box \rightarrow p$ is *necessarily* equivalent to $(\alpha \wedge p) \Box \rightarrow p$. Thus, $\alpha \Box \rightarrow p$ is necessarily equivalent to a logical truth, so $\alpha \Box \rightarrow p$ is itself necessary. So, as in the simple triviality argument, it is necessarily the case

⁴ The uniqueness of closest situations is convenient here, but nothing rests on it.

that $\alpha \Box \rightarrow p$ is true in all epistemically accessible worlds; so knowing $\alpha \Box \rightarrow p$ is trivial. The move from (AK) to $(\alpha K')$ does not seem to have improved things. By a very similar argument, Fara (2010) also arrives at this conclusion.

Before I turn to my own solution to this quandary, it is worth considering Edgington's (2010) own rebuttal. She responds that it is not *prima facie* impossible to have non-trivial knowledge of actuality in a non-actual situations, since to her the "received wisdom about the causal and spatio-temporal isolation of counterfactual situations is overdone" (p. 48). She argues that it is quite plausible that some situations *share a causal history* with the actual situation up to some "point of departure" (p. 51) and, she continues, that someone in a counterfactual situation can specify the actual situation by referring to this shared history and the point of departure. An example where this makes sense is (Comet): someone in the non-actual situation that is specified by *the experiment succeeded* can consider the experiment as a point of departure, and specify actuality by referencing the experiment and its possible outcomes: "We will think, about the other possibility, that if the experiment had gone wrong, p would never have been known." (Edgington 2010, p. 48).

That is, Edgington proposes to meet Williamson's challenge by specifying non-actual situations by appealing to a branching event, e.g. by saying *in the other outcome (of some event)* (p. 42). But this is not yet precise enough. If the event has more than two possible outcomes, it is unclear how to refer to the actual situation: if I roll a die, there are six possible situations; if I roll two dice, 36—how do I pick out the actual situation, if there is no unique *other*? (This concern is also raised by Fara 2010). Edgington (2010) contends that one can *directly* refer to some situations by referring to *another outcome of a branching event*. But if there are multiple other outcomes, one must somehow distinguish the actual outcome. Supposedly, one distinguishes it by a description α (say, *the outcome in which...*).

A variant of Williamson's argument shows that even when appealing to shared histories, this α must be such that knowing $\alpha \Box \rightarrow p$ is trivial. The present suggestion is that one can know about other situations by knowing that they share a causal history, where their history branches off, and referring to the branching point. Again, α must be specific enough to distinguish the unique actual situation. So for all r in the actual situation, in all situations sharing this history and branching point, $(\alpha \wedge p \Box \rightarrow r) \leftrightarrow (\alpha \Box \rightarrow r)$, lest knowledge of α fail to be knowledge of actuality. Therefore, while this biconditional is not outright necessary (it may fail in causally isolated situations), it is a necessary consequence of sharing the history and branching point. But since we suppose that in these situations it is known that they share a history and branching point, all from them epistemically accessible worlds share this history too, and hence satisfy the biconditional. So knowing that $\alpha \Box \rightarrow p$ again requires no more than knowing a triviality.

Thus, Williamson's challenge remains forceful, even given Edgington's appeal to shared histories. I argue that the objection can nevertheless be met. In what follows, I propose a suitable clarification of $(\alpha K')$ that follows many of Edgington's intuitions, but without pretending that this is what she might mean or should mean. My account replaces \diamond in $(\alpha K')$ with another counterfactual conditional. This allows me to give a general definition of what the specification α is and to define in *which particular* worlds

knowledge of $\alpha \Box \rightarrow p$ counts as knowledge of actuality. Williamson's objection is then avoided by denying that knowledge of α must necessarily be knowledge of actuality.

2.2 The counterfactual pursuit of inquiries

For the sake of clarity, I return to talk of possible *worlds* instead of situations, but my proposal could be executed in a situation semantics as well. Now, my discussion of the knowability principle has been so far silent on the *means* by which we gain knowledge. As the examples (Oats) and (Comet) suggest, the anti-realist should not be interested in non-actual worlds (or situations) where it just so *happens* that someone knows $\alpha \Box \rightarrow p$. Rather, the relevant worlds are those that differ from the actual world in that someone has *done* something to get that knowledge (experimenting, counting, measuring, ...).

Thus, instead of appealing to the unclear notion of a shared history, I suggest that the worlds where knowledge of α constitutes knowledge of the actual world are those worlds that differ from the actual world in the successful performance of some *action* that brings about the requisite knowledge. In those worlds, someone might know that (*even*) if I hadn't done such-and-such, it would (still) be true that this-and-that. In (Oats), in a counterfactual situation where I counted, I know that *even if I hadn't counted, I would still have eaten N oats*. Since I actually did not count, knowledge of this counterfactual constitutes knowledge of actuality (and I argue that this is *non-trivial* knowledge). This plan can be spelled out formally in the language of counterfactual conditionals.

For the purposes of a formal statement, we can abstractly consider *all possible courses of inquiry*. Write *Inq* for the set of these courses. For some $i \in \text{Inq}$, write $sp(i)$ for the proposition that i has been successfully pursued (at some time). That is, $w \Vdash sp(i)$ iff i is pursued in w at some time in a way that successfully imparts knowledge. For example, some $i \in \text{Inq}$ may denote some measuring procedure of some object and $sp(i)$ expresses that the procedure concluded as planned, which imparts knowledge of the object's measurements.⁵

Then I can state (ART) as a precise formulation of the anti-realist thesis. Note that (ART) does not contain any actuality operators A . I will talk about (ART) as being evaluated at the actual world, but nothing would stop one from evaluating (ART) at arbitrary worlds.

$$\text{(ART)} \ p \text{ iff } Kp \vee \exists i \in \text{Inq} (\neg sp(i) \wedge (sp(i) \Box \rightarrow K(\neg sp(i) \Box \rightarrow p))).$$

This is a bit of a mouthful, so let me spell it out: p is actually true if and only if either (i) p is actually known (at some time), or (ii) it is metaphysically possible to know that actually p , which means that there is a course of inquiry i that is not actually

⁵ The idea to ground knowability in the possible pursuits of knowledge-inducing courses of inquiry has been anticipated by Artemov and Protopopescu (2013) and Fischer (2013). My proposal is different in that I consider a counterfactual conditional in the worlds that result from these possible pursuits. This allows me to retain factiveness, which Artemov and Protopopescu and Fischer must discard (see an earlier footnote). Their proposals also only make some truths knowable (those they call *stable truths*), whereas my proposal allows for the knowability of all actual truths, including Fitch propositions.

successfully pursued, but if successfully pursued would lead to the knowledge that hadn't it been pursued, p . Roughly, (ART) arises from $(\alpha K')$ by specifying α to be $\neg sp(i)$, and replacing \diamond by $sp(i) \Box \rightarrow _$ for some suitable i .

(ART) can be understood as a claim about how extensive our means to obtain knowledge are. The anti-realist claims that these means are as extensive as truth itself. A realist, then, may claim that for some truths, there are no means that impart (even counterfactual) knowledge about them. However, the realist must now produce a more substantial case than appealing to the Fitch proposition that p and p is not known.

As seen in the example (Comet), there is an intuitive sense in which Fitch propositions are knowable. (ART) makes this sense precise. If $p \wedge \neg Kp$ and (ART) is true, there is a course of inquiry i and a possible world w such that $w \Vdash sp(i) \wedge K(\neg sp(i) \Box \rightarrow (p \wedge \neg Kp))$ (a world in which, say, an experiment succeeded that actually did not succeed). Nothing paradoxical follows from this, since from the successful pursuit of i or from the knowledge of the counterfactual it does not follow that $w \Vdash K(p \wedge \neg Kp)$.

One may wonder what kind of i is suitable to witnessing the knowability of a truth p . I cannot give a general recipe. For many p , a suitable i is a procedure that is known to determine whether p without changing the truth value of p . If one pursues such an i for a true p , then one can counterfactually know that (even) if i hadn't been pursued, p would (still) be true precisely because pursuing i imparted the knowledge that p , and one moreover knows that pursuing i did not change the truth of p .

However, this is not generally the case. Sometimes, pursuing some i imparts the knowledge that $\neg sp(i) \Box \rightarrow p$ without imparting the knowledge that p . The most salient example in this discussion are Fitch propositions $p = (q \wedge \neg Kq)$: one can know $\neg sp(i) \Box \rightarrow (q \wedge \neg Kq)$, but one cannot come to this knowledge by learning that $q \wedge \neg Kq$, since it is paradoxical to know this. But this does not conflict with (ART). As in (Comet), the knowability of a Fitch proposition $q \wedge \neg Kq$ can be witnessed by a course of inquiry i that imparts the knowledge that q without changing the truth value of q , plus the background knowledge that pursuing i is the only way to know q . Then, if one successfully pursues i , one can learn that even if i hadn't been pursued, q would obtain and moreover that q would not be known (since i is the only way to learn q).

There are also K -free examples of propositions p where one can come to know that $\neg sp(i) \Box \rightarrow p$ without knowing that p . One salient class are propositions denoting properties susceptible to an *observer effect*, i.e. a property that changes when one measures it. An example is (Hydrogen).

(Hydrogen). Suppose I have a container that contains either helium or hydrogen. It is actually true that $p =$ *the container contains hydrogen*, but this is not known. Further suppose that any course of inquiry that could tell anyone anything about the container's contents would involve opening it. In any possible world where someone does so, the gas escapes, leaving the container empty.

In (Hydrogen), any world where I make an attempt to determine whether p is a world in which p is false (since the gas escaped). So in these worlds p is not known. In worlds where p is true, it is not known that p , since any inquiry that does not involve opening the container fails to impart knowledge about p . So there is no course of inquiry i that imparts the knowledge that p . Formally, $p \wedge \neg \exists i \in Inq : sp(i) \Box \rightarrow Kp$.

Observer effects are counterexamples to the verificationist thesis that *for every truth, there is a course of inquiry that verifies it*. However, the failure of the verificationist thesis is compatible with the anti-realist thesis (ART), which only requires that if p is actually true, then there is some $i \in Inq$ such that for the closest world w where $sp(i)$, $w \Vdash K(\neg sp(i) \Box \rightarrow p)$. It is not required that p is known at w or even that p is true at w . For (Hydrogen), such an i would be *to open the container and hold a flame to the escaping gas*. Say that in the closest world where I do so, the gas will light, which imparts the knowledge that it is hydrogen (and not helium). Hence, in this world I know the *counterfactual* that *had I not opened the container, it would contain hydrogen*. I do not, however, know that *the container contains hydrogen*, since this proposition is false at that world.

There is a host of other cases that are similar to (Hydrogen), e.g. measuring voltage (adding a voltmeter to a conductor changes the resistance, hence changes voltage) or measuring air pressure in a tire (some air must escape into the measuring apparatus, changing the pressure of the tire). None of them are in obvious conflict with (ART).⁶ Suitable inquiries for these cases might involve, say, mathematically correcting for the distortion introduced by measurement.

In any case, the claim that (ART) is true means that even for propositions p susceptible to observer effects, there are courses of inquiry that allow one to determine that p was the case before the inquiry was pursued, even if the inquiry necessarily involves changing the state. A suitable inquiry needs to be determined on a case by case basis—(ART) merely claims that there *is* one. My goal here is not to defend the truth of (ART), but only the coherence of (ART) *vis-à-vis* the gamut of triviality objections.

2.3 How (ART) avoids the triviality objections

First, one can demonstrate that the right-hand-side of (ART) expresses a factive notion of knowability. Note the following plausible assumption about the semantics of counterfactual conditionals. For arbitrary formulae φ and arbitrary worlds x and y with $x \Vdash \varphi$ and $y \Vdash \neg\varphi$: if x is the closest φ -world to y , then y is the closest $\neg\varphi$ -world to x . Call this property Symmetry. Now, recall (ART).

$$\text{(ART)} \quad p \text{ iff } Kp \vee \exists i \in Inq(\neg sp(i) \wedge (sp(i) \Box \rightarrow K(\neg sp(i) \Box \rightarrow p))).$$

Take a proposition p for which the right-hand side of (ART) is true. If Kp then p is actually true. Otherwise, there is an i such that for the $sp(i)$ -world w that is closest to the actual world, $w \Vdash K(\neg sp(i) \Box \rightarrow p)$. Since knowledge is factive, $w \Vdash (\neg sp(i) \Box \rightarrow p)$. By Symmetry, the closest $\neg sp(i)$ -world to w is the actual world. Thus, p is true in the actual world. Hence, knowing in the closest $sp(i)$ -world that $\neg sp(i) \Box \rightarrow p$ is knowing something about the actual world. In particular, no falsities are knowable.

⁶ Note that uncertainty principles in quantum mechanics are *not* observer effects, but describe fundamental properties of quantum systems. It is true that Heisenberg's principle, if correct, entails that one cannot know, not even counterfactually, that *this particle has position x and momentum y* for x and y below the threshold of possible precision of measurement. But this may just mean that there is no true proposition of the form *this particle has position x and momentum y* . See Putnam (1981, 1982) for discussion.

That is, like Edgington (1985), I propose that someone in a non-actual world (situation) can know that *actually* p by having counterfactual knowledge about a description α . But I make two clarifications: α is now clearly defined to be $\neg sp(i)$ for some i that depends on p ; and knowledge about α does not constitute knowledge of actuality in every world (but only in the closest $sp(i)$ -world).

Hence, knowledge about $\alpha = \neg sp(i)$ is non-trivial, because in worlds that are not $sp(i)$ -closest to the actual world, knowledge about α is not knowledge of actuality. In particular, $\alpha \Box \rightarrow p$ might fail at those worlds. Hence it is not the case that $\alpha \Box \rightarrow p \leftrightarrow \Box(\alpha \Box \rightarrow p)$ and so (ART) avoids the simple triviality argument.

Similarly, (ART) also avoids the Williamsonian version of the triviality objection (which was a worry that Edgington (2010) left open). Let p be a contingent actually true proposition whose knowability is witnessed by some course of inquiry i . We can assume that $\neg p$ and $\neg sp(i)$ are compossible (but see below). Let w be a witness for $\Diamond(\neg p \wedge \neg sp(i))$. Since w satisfies neither p nor $sp(i)$, it is false at w that $\neg sp(i) \Box \rightarrow p$. But of course $(p \wedge \neg sp(i)) \Box \rightarrow p$ is a logical truth and hence true at w . Thus, the following biconditional is false at w :

$$(\neg sp(i) \Box \rightarrow p) \leftrightarrow ((p \wedge \neg sp(i)) \Box \rightarrow p).$$

But this means that this biconditional is not necessary. Similarly, it is not in conflict with (ART) to suppose that the biconditional is not necessarily known: since its right-hand-side is a logical truth, the biconditional is known whenever $\neg sp(i) \Box \rightarrow p$ known. But $\neg sp(i) \Box \rightarrow p$ may be false at some worlds, so to know it is non-trivial.

Of course, in the world that witnesses the knowability of p (i.e. the $sp(i)$ -world closest to the actual world), $\neg sp(i) \Box \rightarrow p$ is true and known, and so the biconditional is true and known. But the *contingent* truth or knowledge of a biconditional does not mean that knowing its one side is equally good as knowing its other side. For comparison, consider that it happens to be the case in the actual world that *grass is green iff* $1 = 1$, since both sides of that biconditional are true. But the fact that knowledge of *grass is green* is substantive knowledge about the actual world is of course not incompatible with knowledge of $1 = 1$ being trivial knowledge. This is because *grass is green iff* $1 = 1$ is neither necessary nor trivial to know.

Similarly, knowing a contingently true counterfactual $q \Box \rightarrow p$ is substantive knowledge about some possibly non-actual world, despite the biconditional $q \Box \rightarrow p \text{ iff } 1 = 1$ being (contingently) true.⁷ In perfect analogy, there are worlds where knowing that $\neg sp(i) \Box \rightarrow p$ is substantive knowledge about actuality, even though in these worlds the counterfactual is contingently equivalent to the triviality that $(p \wedge \neg sp(i)) \Box \rightarrow p$.

The assumption that $\neg p$ and $\neg sp(i)$ are compossible hides a final difficulty, and pressing this difficulty seems to reveal another kind of triviality. Take any $i \in Inq$ and let $p = \neg sp(i)$. Obviously, $\neg p$ and $\neg sp(i)$ are not compossible. So the above biconditional is necessary for these i and p . Then, indeed, it is trivial that $sp(i) \Box \rightarrow K(\neg sp(i) \Box \rightarrow p)$ because $\neg sp(i) \Box \rightarrow \neg sp(i)$ is a logical truth. So (ART) trivialises knowability of which courses of inquiry are actually pursued. But this does not seem to

⁷ Some may argue that knowledge of $q \Box \rightarrow p$ is not knowledge about some other world, but knowledge about the causal structure of the actual world. Those may consider my talk of possible worlds as mere convenience, and interpret (ART) as a statement about the causal structure of the actual world.

be a trivial kind of knowledge. More generally, (ART) trivialises the knowability of all actual truths that logically follow from not having performed some course of inquiry.

But these cases are a distraction. The anti-realist can simply rewrite (ART) to *demand* that the pursuit of the i that witnesses the knowability of a proposition p be independent of p . As said above, (ART) should be read as a thesis about how extensive our means to gain knowledge are. To say that pursuing these means ought to be independent of the proposition whose knowability they witness simply amounts to a slightly stronger thesis. To do so formally, one can revise (ART) to (ART'): (I have underlined the added clause for visibility):

$$(ART') \quad p \leftrightarrow Kp \vee \exists i \in Inq(\neg sp(i) \wedge \neg \Box K(\neg sp(i) \rightarrow p) \wedge (sp(i) \Box \rightarrow K(\neg sp(i) \Box \rightarrow p))).$$

For those propositions p that necessarily are known to follow from $\neg sp(i)$ for some $i \in Inq$, the refined principle (ART') demands that p 's knowability be witnessed by another $j \in Inq$. This demand rules out trivialising knowledge about which i are pursued: to know that some i is not pursued, one needs to pursue some j that is not i .

There is a formal point here about necessary propositions p that are never actually known (say, that some outrageously large number is prime, *cf.* Tennant 1997). To say that such a p is knowable should be to say that there is some i (say, a process of factorisation) such that $sp(i) \Box \rightarrow K(\neg sp(i) \Box \rightarrow p)$. But this does not follow from (ART'), since if p is a logical truth, then $\Box K(\neg sp(i) \rightarrow p)$ is vacuously true for any i . However, this point is moot as long as we are operating within the idealised modal framework where knowledge is closed under inference (hence where all logical truths are known). In the present framework, logical truths are not counterexamples to (ART'), since if p is a logical truth, Kp is true, so the right-hand-side of (ART') is true. If we move to a framework where not all logical truths are known, then $\Box K(\neg sp(i) \rightarrow p)$ can fail for some i , and (ART') entails—as it should—that for some such i , $sp(i) \Box \rightarrow K(\neg sp(i) \Box \rightarrow p)$

I should note that, despite appearances, moving from (ART) to (ART') does not put me in the idle business of chasing exceptions and adding stop-gap fixes. Rather, I am still offering increasingly precise formalisations of a single claim. That claim is now recognisable to be (*).

- (*) Given some actually true p , there is a course of inquiry i that is not successfully pursued, but were it successfully pursued, would impart the knowledge that had it not been (i.e. in actuality), p .

All examples and arguments I have presented so far, including those that motivated (AK), also motivate and defend (*). Aside from (Oats) and (Comet), mundane examples for (*) are abound, e.g. the procedure i to *look at the ceiling* which, if successfully pursued, would impart the knowledge that *hadn't I looked, there (still) would be a fly on the ceiling* (Edgington 1985). An anti-realist following me on characterising knowability as in (ART) claims that such i 's can be found for *all* actual truths p (which would mean that all actual truths are knowable, in my sense of knowability).

In (ART) I have formalised the 'would impart the knowledge' of (*) simply as a counterfactual conditional. As seen, this trivialises the knowability of which inquiries are pursued. I find that (*) loses no plausibility when I spell out 'would impart the

knowledge' to include that the successful pursuit of i must be *relevant* to the knowledge that *hadn't i been successfully pursued, p* . All the same examples still apply and, in fact, the move to demand relevance only seems to sharpen and clarify what is meant by 'impart' in (*).

To give proper due to this, one could introduce the concept of a counterfactual relevant conditional and use it to phrase (ART). But sticking to the usual counterfactual semantics, one can approximate such a conditional by moving from (ART) to (ART'). Observe that whenever $\neg sp(i) \rightarrow p$ is necessarily known, then successfully pursuing i is *not* relevant to the knowledge that $\neg sp(i) \Box \rightarrow p$. These are exactly the cases that are excluded by the added clause in (ART'). So (ART') derives from (ART) not as a stop-gap measure, but as a more precise way to spell out (*).

Finally, an anonymous reviewer points out to me another kind of triviality argument that has recently been developed by Heylen (2018). He argues that even if Edgington's principle (AK), repeated below, is interpreted in a suitable semantics for actuality that avoids the derivations of *every truth is known*, it still follows that *possibly, every truth is known* which is equally unwelcome (or so Heylen argues).⁸ The argument goes as follows.

$$(AK) Ap \rightarrow \Diamond KAp.$$

To avoid the simple triviality argument from $Ap \leftrightarrow \Box Ap$, Heylen follows Rabinowicz and Segerberg (1994) in interpreting Edgington's principle (AK) in a semantics where worlds are indexed with different *reference worlds*. One evaluates propositions at a world as usual, except that Ap is true at a world iff p is true at that world's reference world. Since what is the reference world may vary across epistemic accessibility, it does not follow from p being true at the actual reference world that Ap is true at all epistemically accessible worlds. Hence Ap does not trivially entail KAp .

However, Heylen (2018, Theorem 2.1) points out, the frame condition for (AK) on such semantics is this: given a reference world v and any metaphysically accessible w , all from w epistemically accessible worlds must have the reference world v . Then, suppose that the actual world is the reference world and that p is actually true. According to (AK) there is a w where KAp . That is, all worlds epistemically accessible to w satisfy Ap . All these worlds share the actual world as reference world by the frame condition. But this means that for any q that is actually true, Aq is true at all worlds epistemically accessible to w . Hence, all actual truths are known at w . So, *possibly, every actual truth is known to be actually true*.

This argument does not apply to (ART'), simply because (ART') replaces the actuality operator A with a counterfactual that only corresponds to actuality in the $sp(i)$ -world closest to the actual world. To wit: If p is actually true, there is an i such that in the $sp(i)$ -world w closest to the actual world, $K(\neg sp(i) \Box \rightarrow p)$ is true. This means that all worlds epistemically accessible to w make $\neg sp(i) \Box \rightarrow p$ true. But this does not mean that the actual world is the $\neg sp(i)$ -world closest to these worlds. Consider that a v that is epistemically accessible from w may differ quite substantially from

⁸ Artemov and Protopopescu (2013) point out that their stable knowability principle entails possible omniscience. They do not see this as a problem, but as "entirely what one would expect from the idea that all truths are knowable" (p. 3370). I will not adjudicate.

w (sharing only those truths that are known at w) and so be distant from w and hence also from the actual world (which is close to w). Therefore, it does not follow that for any other actually true q it is the case that all these worlds make $\neg sp(i) \Box \rightarrow q$ true.

There is an entirely technical point here. Abstracting somewhat, Heylen makes a general objection to factive notions of knowability: for a knowability principle to be factive, its frame condition must be such that the worlds witnessing the knowability of some p are all accessing or referencing the actual world (lest actual falsities be knowable). But if this is the case, then these worlds witness the knowability of *all* actual p , entailing that *possibly, all truths are known*. Such an argument cannot work for (ART'), since (ART') replaces modal operators (such as \Diamond or A) with counterfactual conditionals. There are no frame conditions for counterfactuals, since—as opposed to \Diamond and A —which world they access depends partly on valuation, not exclusively on the frame.

3 Summary and conclusion

As is known since Edgington's landmark 1985 paper, Fitch's paradox does not arise when the anti-realist principle is taken to be about *counterfactual* knowledge of *actual* truth. This reveals the paradox to be a misunderstanding: due to the imprecisions of casual talk about truth and actuality, *truth is what is knowable* has not been recognised for the claim that it should be: *actual truth is what is possible to know to be actual*.

But this explication of the anti-realist principle requires that there be non-trivial counterfactual knowledge of actual truth. Williamson's (1987) forceful objection suggests that no such thing can exist. However, the solution (ART') developed in the foregoing discussion shows that this is not so.

Let Inq be a set of possible courses of inquiry, and for any $i \in Inq$ let $sp(i)$ be the proposition stating that i is successfully pursued.

$$(ART') \quad p \leftrightarrow Kp \vee \exists i \in Inq (\neg sp(i) \wedge \neg \Box K (\neg sp(i) \rightarrow p) \wedge (sp(i) \Box \rightarrow K (\neg sp(i) \Box \rightarrow p))).$$

(ART') can be considered a formalisation of the informal paraphrase (*), which itself is an explication of the knowability principle.

- (*) Given some actually true p , there is a course of inquiry i that is not successfully pursued, but were it successfully pursued, would impart the knowledge that had it not been (i.e. in actuality), p .

Williamson's triviality objection is deflected by appealing to a restricted class of metaphysically possible worlds: those that diverge from the actual world in the counterfactual pursuit of some inquiry i . In such worlds, it is possible to have non-trivial counterfactual knowledge of the actual world by being acquainted with a description of the actual world, namely $\alpha = \neg sp(i)$. This explanation of counterfactual knowledge addresses Williamson's general objection that counterfactual knowledge of actuality is philosophically unsound: it requires no more than there being true counterfactuals and there being knowledge of counterfactuals, which puts it on philosophically steady

ground. On the technical side, Williamson's argument is avoided by the observation that the trivialising biconditional is not a metaphysical necessity.

In conclusion, the anti-realist has at her disposal a version of the thesis that *truth is what is knowable* that is metaphysically sound and clarifies the relevant notions so that Fitch's paradox disappears. But this just means that her thesis stands up to the Fitch objection—not that her thesis is true. I happen to think that it is not. However, the right-hand-side of (ART') expresses a notion of knowability that is useful even if (ART') itself is false. By appealing to the counterfactual pursuit of inquiries, we find a notion of knowability that is factive and on which mundane facts that intuitively seem knowable—such as there being *N* oats in my bowl—are in fact knowable. I leave open here whether all—and if not, how many—facts are knowable in this sense.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Artemov, S., & Protopopescu, T. (2013). Discovering knowability: A semantic analysis. *Synthese*, *190*(16), 3349–3376.
- Beall, J. C. (2000). Fitch's proof, verificationism, and the knower paradox. *Australasian Journal of Philosophy*, *78*(2), 241–247.
- Brogaard, B., & Salerno, J. (2013). Fitch's paradox of knowability. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*, winter (2013th ed.). Stanford: Metaphysics Research Lab, Stanford University.
- Dummett, M. A. (1991). *The logical basis of metaphysics*. Cambridge: Harvard University Press.
- Edgington, D. (1985). The paradox of knowability. *Mind*, *94*(376), 557–568.
- Edgington, D. (2010). Possible knowledge of unknown truth. *Synthese*, *173*(1), 41–52.
- Fara, M. (2010). Knowability and the capacity to know. *Synthese*, *173*(1), 53–73.
- Fischer, M. (2013). Some remarks on restricting the knowability principle. *Synthese*, *190*(1), 63–88.
- Fuhrmann, A. (2014). Knowability as potential knowledge. *Synthese*, *191*(7), 1627–1648.
- Heylen, J. (2018). Factive knowability and the problem of possible omniscience. *Philosophical Studies*. <https://doi.org/10.1007/s11098-018-1180-x>.
- Kvanvig, J. (1995). The knowability paradox and the prospects for anti-realism. *Noûs*, *29*(4), 481–500.
- Putnam, H. (1981). Quantum mechanics and the observer. *Erkenntnis*, *16*(2), 193–219.
- Putnam, H. (1982). Why there isn't a ready-made world. *Synthese*, *51*(2), 141–167.
- Rabinowicz, W., & Segerberg, K. (1994). Actual truth, possible knowledge. *Topoi*, *13*(2), 101–115.
- Rückert, H. (2004). A solution to fitch's paradox of knowability. In S. Rahman & J. Symons (Eds.), *Logic, epistemology, and the unity of science* (pp. 351–380). Berlin: Springer.
- Tennant, N. (1997). *The taming of the true*. Oxford: Oxford University Press.
- Tennant, N. (2001). Is every truth knowable? Reply to Williamson. *Ratio*, *14*(3), 263–280.
- Tennant, N. (2002). Victor vanquished. *Analysis*, *62*(2), 135–142.
- Tennant, N. (2009). Revamping the restriction strategy. In J. Salerno (Ed.), *All truths are known: New essays on the knowability paradox*. Oxford: Oxford University Press.
- Tennant, N. (2010). Williamsons woes. *Synthese*, *173*(1), 9–23.
- Williamson, T. (1987). On the paradox of knowability. *Mind*, *96*(382), 256–261.
- Williamson, T. (2000). Tennant on knowable truth. *Ratio*, *13*(2), 99–114.
- Williamson, T. (2009). Tennant's troubles. In J. Salerno (Ed.), *Every truth is known: New essays on knowability* (pp. 183–204). Oxford: Oxford University Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.