KANT, THE PARADOX OF KNOWABILITY, AND THE MEANING OF 'EXPERIENCE'

Forthcoming in *Philosopher's Imprint*

Andrew Stephenson

Introduction

It is often claimed that anti-realism is a form of transcendental idealism or that Kant is an anti-realist.[1] It is also often claimed that anti-realists are committed to some form of knowability principle to the effect that all truths (or at least all truths of a certain class) are knowable and that such principles have problematic consequences.[2] It is therefore natural to ask whether Kant was committed to any such principle, and if he was, whether this leads him into similar difficulties. Both transcendental idealism and anti-realism aim to provide a middle way between realism and idealism. A logical proof published by Frederic Fitch in 1963 (though first conveyed to him by Alonzo Church in 1945) appears to show that anti-realism fails in its aim because it collapses into idealism. Can a related proof show that transcendental idealism collapses in the same way? I argue that, initial appearances to the contrary, it cannot.

---

[1] See Strawson (1966, 16), Putnam (1981, 60ff.; 1987, 41ff.), Posy (1983; 1984), Walker (1983; 1995), Stevenson (1983), Hanna (1993), Moran (2000), Baldwin (2001), Allais (2003; 2015, 207-31), and Moore (2012, 362-6).

[2] For an overview and some recent contributions to the debate, see Salerno (2009).

The paper is in two parts. In the first part I set up the problem and in the second part I solve it.

In §1.1 I present evidence that suggests Kant is indeed committed to a knowability principle and I show that a Fitch-Church style proof can be constructed on this basis. Kant does not think that all truths whatsoever are knowable, but it can seem as though he is committed to the claim that all *empirical* truths are knowable, and on moderate background assumptions this entails that no empirical truth is unknown. In §1.2 I show that with a few additional assumptions we can also prove that all a priori truths are knowable and that no a priori truth is unknown. This is an interesting result with more general philosophical lessons concerning how certain classes of truth relate within a framework of knowability. But it is a little unfair to Kant. Arguably we ought to further restrict our candidate Kantian knowability principle to what I call *purely* empirical truths, and doing so blocks the seepage into the a priori realm. However, this move would still leave Kant forced to concede that there are no unknown purely empirical truths, which is hardly more palatable.

Thus in the second part of the paper I explore an alternative route. The evidence for Kantian knowability relies on interpreting Kantian experience as a form of knowledge. This is a standard view but it is not always correct. Sometimes Kantian experience is something more like final science. In §2.1 I explain this conception of experience and apply it to the case at hand. Because, for Kant, experience so conceived is an unachievable epistemic *ideal*, it expresses no knowability principle to define truth in terms of it. Arguably, however, this proposal would still leave Kant committed to the claim that all purely empirical truths can be the objects of justified belief, and it has been objected that this kind of principle remains just as susceptible to Fitch-Church style reasoning.[3] In §2.2 I argue that Kant has exactly the resources needed to rebut such an objection.

Kant's theory of truth has both realist and idealist aspects and is in a way anti-realist. But Fitch-Church style reasoning alone cannot show us that the theory is absurd.

---

[3] See Mackie (1980, 91-2), Edgington (1985, 558-9), Tennant (1997, 252-9), and Wright (2000, 355-7).

Kant famously denies that we can have knowledge of things as they are in themselves. All we can know about, according to the doctrine of transcendental idealism, is how things appear to us; all our knowledge is knowledge of appearances. Yet at the same time Kant seems to concede that there is a way that things in themselves are and thus that there are truths about things in themselves. He says: 'knowledge reaches appearances only, leaving the thing in itself as something actual for itself but unknown by us' (Bxx).[4] If this is right, then Kant does not think that all truths whatsoever are knowable and so would not sign up to a wholly unrestricted knowability principle. But the matter is controversial. Let us put it aside and concentrate on appearances. Does Kant think there are truths about appearances that must in principle be knowable?

Kant divides knowledge into two kinds: a priori and empirical. A priori knowledge is 'knowledge absolutely independent of all experience' while empirical knowledge 'has its sources a posteriori, namely in experience' (B2-3). For knowledge to have its sources in experience and thus fail to be absolutely independent of all experience, it is not enough that it merely be connected to experience in some way. After all, for Kant, there is a sense in which *all* our knowledge begins with experience (B1). The requirement is rather that empirical knowledge, unlike a priori knowledge, involves appeal to particularities in the information provided through the senses, beyond whatever was required to grasp the meaning of the constituent terms. Kant might call such particularities the 'matter' of empirical intuition. This gloss is still very rough, but it will suffice for present purposes.

Kant often attaches the labels 'a priori' and 'empirical' directly to propositions, or judgements, and we can safely assume that they may also be applied directly to truths.[5] Where 'independently of experience' and 'with the aid of experience' are to be understood in the sense just outlined: if a truth is knowable independently of experience, then it is an a priori truth; if

---

[4] Cf. B332, A366, B641-2. With the exception of the *Critique of Pure* Reason, which is cited using the traditional A/B format, all references to Kant's works are given by the volume and page number in the German Academy edition and are accompanied by a short English title. The details of the translations I have consulted are given in the list of references at the end.

[5] He uses the phrase 'empirical truth' himself on several occasions, for example at A191/B236, A202/B247, A451/B479, A492/B520, and A651/B679.

a truth is knowable only with the aid of experience, then it is an empirical truth. Note that this is not yet a commitment to any form of knowability. As Kant sets up these distinctions, the path to knowledge is *sufficient* to determine the kind of truth known. It is the converse of these claims that would give us the kind of epistemic theory of truth that underlies (semantic) anti-realism and knowability principles. Let us begin with empirical truth. Our question is whether Kant ever *defines* empirical truth in terms of knowability.

Here is one passage in which Kant can very plausibly be read as doing exactly that:

> That there could be inhabitants of the moon, even though no human being has ever perceived them, must of course be admitted; but this means only that in the possible progress of experience we could encounter them; for everything is actual that stands in one context with a perception in accordance with the laws of the empirical progression. Thus they are real when they stand in an empirical connection with my real consciousness, although they are not therefore real in themselves, i.e., outside this progress of experience… To call an appearance a real thing prior to perception means either that in the continuation of experience we must encounter such a perception, or it has no meaning at all. (A493/B521; cf. A218-26/B266-73, A155-6/B194-5)

Suppose that there are inhabitants of the moon. This would be an empirical truth. To come to know it, we might have to explore the moon, appealing to particularities in the information provided through the senses. It could never suffice to appeal only to very general features shared by any information whatsoever that comes to us through the senses, such as its spatiotemporal form, just as it could never suffice to appeal to pure reason alone. Thus what Kant seems to be doing here is defining at least empirical truth in terms of possible experience. This certainly looks like a form of anti-realism. And once we note that Kant is standardly taken to define experience in turn as a form of *knowledge*, it seems more specifically that he is expressing a direct commitment to the claim that all empirical truths are

knowable.[6] For then he seems to be saying that it is *just what it is* to be an empirical truth to be an object of possible knowledge. I will return to this and related passages in more detail in §2. For now let us assume that this proposed, standard reading is correct. It remains a substantial question whether such a restricted knowability principle can be used to run a version of the familiar Fitch-Church proof. In fact it can.

In addition to the standard modal and epistemic operators – ◊ for 'it is possible that', □ for 'it is necessary that', and K for 'it is known by someone at some time that' – we make use of an empirical truth operator E for 'it is an empirical truth that'. We can then express our candidate knowability principle as an axiom schema:

(KPE) $E\phi \rightarrow \Diamond K\phi$

The kind of modality involved in anti-realist knowability principles is somewhat open. Here it is taken to be metaphysical. It is not entirely clear which correlate this contemporary notion has in Kant's framework. It may be that it has no single, precise correlate. Kant distinguishes various kinds of modality and it is a matter of controversy what they amount to and how they relate.[7] Certainly the kind of possibility involved in Kant's talk of possible experience is more restrictive than mere logical possibility. It is some form of what he calls 'real' possibility (Bxxvi, A218-35/B265-87, A596/B624). Thus it is assumed here that real possibility entails metaphysical possibility, though I stay neutral on the converse and, more generally, assume very little about the nature of the modality in play. Relatedly, the quantification over persons implicit in K should be understood throughout as restricted to humans, or at least to beings with intellectual and sensible forms identical to our own. KPE says that if ϕ is an

---

[6] See, e.g., B147, B165-6, B218, B234, B277; *Prolegomena* (4:302). The word Kant uses to define experience in these passages is '*Erkenntnis*', which is now generally translated with 'cognition' rather than 'knowledge'. But this is largely for linguistic reasons and the view that Kantian *Erkenntnis* is a form of knowledge remains extremely widespread – see §2.

I go on to call the reading on which Kant thinks that all empirical truths are knowable 'standard' because it is implied by the standard reading of experience, not because it is standard to explicitly attribute the view to Kant (though see Ameriks (1982, 45, 51) and Hanna (2006, 32-3)). Despite common knowledge of the connection between Kant and anti-realism on the one hand and anti-realism and knowability on the other, the current topic has not, to my knowledge, been explored previously.

[7] See Stang (forthcoming) for by far the most comprehensive study to date.

empirical truth, then it is metaphysically possible for beings like us to know ϕ.

We assume the necessitation rule and that the modal operators can be exchanged in the normal way, and we also assume that knowledge distributes over conjunction and is factive:

(K-DIST) $K(\phi\wedge\psi)\rightarrow K\phi\wedge K\psi$

(K-FACT) $K\phi\rightarrow\phi$

Finally, we assume the following principles governing the empirical truth operator E, which will be discussed below:

(EK) $E\phi\wedge\neg K\phi\rightarrow E\neg K\phi$

(E-CLOS) $E\phi\wedge E\psi\rightarrow E(\phi\wedge\psi)$

We are now in a position to construct a Fitch-Church style proof. Central to such proofs is the so-called Moore proposition, $p\wedge\neg Kp$. The first step is to prove that such a proposition is unknowable. We proceed by reductio:

(1) $K(p\wedge\neg Kp)$ assumption

(2) $Kp\wedge K\neg Kp$ 1, K-DIST

(3) $Kp\wedge\neg Kp$ 2, K-FACT on right conjunct

(4) $\neg K(p\wedge\neg Kp)$ 1, 3, discharging assumption

(5) $\Box\neg K(p\wedge\neg Kp)$ 4, necessitation

(6) $\neg\Diamond K(p\wedge\neg Kp)$ 5, modal operator exchange

Now we can prove the main result using KPE and the principles governing E. Again we proceed by reductio, but this time we take as our assumption the claim that some empirical truth is unknown. Call this claim empirical humility. The second step:

(7) $\exists p\ (Ep\wedge\neg Kp)$ assumption

(8) $Ep\wedge\neg Kp$ 7, existential instantiation

(9) $Ep\wedge E\neg Kp$ 8, EK to yield right conjunct

(10) $E(p\wedge\neg Kp)$ 9, E-CLOS

(11) $\Diamond K(p \wedge \neg Kp)$ 10, KPE

(12) $\neg \exists p \ (Ep \wedge \neg Kp)$ 6, 7, 11, discharging assumption

Plausibly, empirical humility – the assumption at (7) – is entailed by Kant's empirical realism. What I mean is that, plausibly, Kant would want his own brand of realism to be of a robust enough nature to allow that there are some unknown empirical truths.[8] If so, and if KPE is in any way integral to transcendental idealism, then Kant's claim to be both a transcendental idealist and an empirical realist looks inconsistent.

There are many ways to respond to Fitch-Church style chains of reasoning. One common strategy is for the anti-realist knowability theorist to adopt an intuitionistic logic. The above version of the proof is valid in intuitionistic logic. But only in classical logic is the negation of empirical humility at (12) equivalent to the omniscience claim that all empirical truths are known: $\forall p (Ep \rightarrow Kp)$. The anti-realist empirical knowability theorist who adopts an intuitionistic logic could consistently deny that all empirical truths are known while accepting, as she seems forced to by the above proof, that there is no unknown empirical truth.[9] Intuitionism certainly displays a Kantian provenance.[10] And to be on the safe side I ensure that the other proofs in this paper are also intuitionistically valid.[11] But I will not explore this avenue here. Arguably the damage has already been done by the time we reach (12). And in any case, I will argue that Kant has a better solution.

A natural place to look at this point would be the principles governing the empirical truth operator E. This will not help either.

---

[8] Even strong phenomenalist readings of Kant tend to concede this much. See, e.g., Van Cleve (1999, 233-5).

[9] See Williamson (1982; 1992).

[10] See especially chapter 2 of Brouwer (1975, 11-101). More recently and on different grounds, Carl Posy (1981; 1983; 1984) has argued that Kant would have adopted an intuitionistic logic for empirical domains only.

[11] I mean intuitionistically valid as regards their base propositional logic. I will occasionally make use of features of the standard inter-definability of the modal operators which those who are also intuitionists about the modal extension of the base logic would reject. I flag these occasions, though it is fairly clear that Kant, following Baumgarten, accepts the standard inter-definability. See, e.g., *Metaphysics L₂* (28:557): 'Necessary is that of which the opposite is impossible'.

The principle EK says that if some empirical truth is unknown, then it must be an empirical truth that that truth is unknown: $E\phi \wedge \neg K\phi \rightarrow E\neg K\phi$. EK would follow from KPE *if* we were to grant that all contingent truths are empirical truths. For if it were necessary rather than contingent that some empirical truth is unknown, KPE would be false. A little more fully, we assume $E\phi \wedge \neg K\phi$ for conditional proof. Applying KPE to the left conjunct gives us $\Diamond K\phi$, which by modal operator exchange is equivalent to $\neg \Box \neg K\phi$.[12] Conjoining this with the right conjunct of our initial assumption yields $\neg K\phi \wedge \neg \Box \neg K\phi$. This tells us that $\neg K\phi$ is a contingent truth, for contingent truths have just this form, i.e. $\psi \wedge \neg \Box \psi$. Thus if all contingent truths are empirical truths we can infer that $\neg K\phi$ is an empirical truth, or $E\neg K\phi$. All that remains is conditional introduction and we have EK.

How secure is the assumption that all contingent truths are empirical truths? Kripke (1980), of course, denies it. One thing to say here would be that we are talking about Kant, and everyone knows that *he* straightforwardly equates a priori truth with necessary truth and empirical truth with contingent truth. I do not want to rely on this for two reasons. First it would threaten to limit the independent philosophical interest of the argument. If empirical knowability only entails a failure of empirical humility under some dubious conflation of metaphysical and epistemic modalities, then perhaps this result is just another example of the troubles to which such a conflation can lead. Second, I think the question of whether Kant conflates the two kinds of modality is far more complicated than is usually assumed.[13] Fortunately, we can avoid it with a little logical manoeuvring.

Note that the above proof of EK did *not* assume that all empirical truths are contingent truths – it said nothing against the necessary a posteriori. The issue was just with the contingent a priori. And the basic point is that neither Kripke's purported cases of such truths, nor any purportedly Kantian cases – perhaps truths concerning our possession of our sensible and intellectual forms – would tell against the claim that it must be an

---

[12] If we wanted to minimize our modal principles, we could instead assume $\Box \neg K\phi$ for reductio and infer $\neg \Diamond K\phi$ by the same restricted (and intuitionistically valid) operator exchange rule employed to derive (6) above – the resulting contradiction would also allow us to infer $\neg \Box \neg K\phi$. (Though see fn.11.)

[13] See Stang (2011) and Divers (1999) for discussion.

empirical truth when some empirical truth is contingently unknown. Formally, what we need to do is weaken our assumption that *all* contingent truths are empirical truths to something neither un-Kripkean nor (potentially) un-Kantian, but which nevertheless allows us to derive EK from KPE.

In the above proof, KPE alone gets us from $E\phi\wedge\neg K\phi$ to $\neg K\phi\wedge\neg\Box\neg K\phi$, which says that $\neg K\phi$ is a contingent truth. It was at this point that we utilized the assumption that all contingent truths are empirical truths, or $\psi\wedge\neg\Box\psi\rightarrow E\psi$. Could we simply utilize $\neg K\phi\wedge\neg\Box\neg K\phi$ directly, thus weakening our assumption to a claim about a proper subset of contingent truths? The following says that all contingent truths *about some proposition being unknown* are empirical truths: $\neg K\phi\wedge\neg\Box\neg K\phi\rightarrow E\neg K\phi$. Formally, this would suffice, and it is a step in the right direction. The putative cases of contingent a priori truths are not truths about propositions being unknown. But it won't quite do. For any contingent a priori truth could still be used to generate a counterexample to this version of our assumption, weaker though it is.

Suppose that there is some contingently true proposition p that can be known a priori. Then presumably it can be known a priori that $\neg\neg p$. One need merely be a priori justified in introducing double-negation. And if it can be known a priori that $\neg\neg p$, then presumably it can also be known a priori that it is not known that $\neg p$. One need merely know a priori that knowledge is factive and be a priori justified in contraposing: $K\neg p\rightarrow\neg p$, so $\neg\neg p\rightarrow\neg K\neg p$.[14] In which case, $\neg K\neg p$ is not an empirical truth, or $\neg E\neg K\neg p$. But p is contingently true, so $\neg\neg p$ is too. Thus without any reason to deny that $\neg p$ could be known if it were true, it would seem contingent that $\neg p$ is unknown, or $\neg K\neg p\wedge\neg\Box\neg K\neg p$. We have an instance of $\neg K\phi\wedge\neg\Box\neg K\phi$ that does not entail $E\neg K\phi$. Those who believe in the contingent a priori would be no happier with the claim that all contingent truths about propositions being unknown are empirical truths than they would be with the claim that all contingent truths simpliciter are empirical truths.

[14] Note that both double-negation introduction and this form of contraposition are intuitionistically valid, though I do not here require them as inference rules, still less that a priori knowledge is closed under known entailment, since I require only that it *can* be known a priori that $\neg K\neg p$ (if it is known a priori that p).

However, we can further weaken our assumption so as to avoid this problem. The possibility of the contingent a priori alone was not enough to generate the above kind of counterexample. It was also crucial that the antecedent in the previous assumption said nothing about whether the unknown proposition in question is true. This is what allowed us to substitute in the contingent a priori falsehood ¬p, which was required for it to be a priori knowable that the proposition was not known. Thus further weakening our assumption by restricting our concern to truths will block such counterexamples. The following says that all contingent truths about *truths* being unknown are empirical truths: $\phi \wedge \neg K\phi \wedge \neg \Box \neg K\phi \rightarrow E \neg K\phi$. Unlike the previous attempt, this further weakening is immune to the above method of generating counterexamples. And it remains the case that none of the putative cases of contingent a priori truths are about truths being unknown. The assumption is secure even in the face of those who believe in the contingent a priori. And the proof of EK from KPE remains straightforward.

We assume $E\phi \wedge \neg K\phi$ for conditional proof and derive $\neg K\phi \wedge \neg \Box \neg K\phi$ in the normal manner. Now, however, we appeal to the factivity of empirical truth, or $E\phi \rightarrow \phi$. Being an empirical truth is a way of being a truth. This allows us to derive ϕ from the left conjunct of our initial assumption. Putting these results together we get $\phi \wedge \neg K\phi \wedge \neg \Box \neg K\phi$, the antecedent of our new, weakened claim that it must be an empirical truth when it is contingently true that some truth is unknown. We then infer $E \neg K\phi$ and conditional introduction yields EK: $E\phi \wedge \neg K\phi \rightarrow E \neg K\phi$. EK follows from KPE independently of any (philosophically and exegetically) dubious conflation of the metaphysical and epistemic modalities.

What about the other principle governing the empirical truth operator? E-CLOS says that conjoining two empirical truths yields an empirical truth: $E\phi \wedge E\psi \rightarrow E(\phi \wedge \psi)$. Again, if we were to go ahead and equate empirical truth with contingent truth, this principle would be straightforward – contingency is closed under conjunction in any normal modal logic.[15]

---

[15] This will be relevant in the next section so it is worth sketching a proof. Assume for reductio that there is some world w in which it is false that $\phi \wedge \neg \Box \phi \wedge \psi \wedge \neg \Box \psi \rightarrow \phi \wedge \psi \wedge \neg \Box (\phi \wedge \psi)$. Then by the truth of the antecedent ϕ and ψ are true in w and there is some world v accessible from w in which $\neg \phi$ and some world u accessible from w in which $\neg \psi$. But if ϕ and ψ are both true in w, the falsity of the

Fortunately, however, the principle is also independently plausible. In a nutshell, empirical truths are those for our knowledge of which experience is necessary, and it is difficult to see how conjoining two such truths could possibly yield either a falsehood or a truth that could be known without appeal to experience.

Nevertheless, one might worry that in the current, Kantian context, E-CLOS is in fact incompatible with KPE. If the conjunction of all empirical truths is a truth about the world of appearances as a whole, then by E-CLOS this conjunction is itself an empirical truth, and so by KPE knowable. But doesn't Kant deny that we can know truths about the world of appearances as a whole? Relatedly, might there not be infinitely many empirical truths? If so, and if they can all be conjoined to form an infinitely complex empirical truth, surely Kant would deny that finite minds like ours could grasp such a thing. A full discussion of the issues these arguments throw up is beyond the scope of this paper. It suffices here to make the following observations.

Take the second problem first. Our language is a traditional, finitary language. If Kant does think there are infinitely many empirical truths, all of which are knowable by someone at some time – though presumably not by the same person at the same time – then this would mean that KPE does not fully capture his commitment to knowability. Nevertheless it would still capture part of his commitment, and the result of the proof, that no member of any finite subset of all empirical truths is unknown, remains highly problematic.

Note that the first problem is now only a problem if Kant thinks there are only finitely many empirical truths. Otherwise E-CLOS, since our language is finitary, does not say that the conjunction of *all* empirical truths is an empirical truth. But suppose that Kant does think there are only finitely many empirical truths, and furthermore that their total conjunction therefore expresses a truth about the world of appearances as a whole (itself

consequent must be due to the falsity of $\neg\Box(\phi\wedge\psi)$, so $\phi\wedge\psi$ must be true in every world accessible from w – contradiction. (It is unclear whether an intuitionist would permit this proof as it effectively relies on either double-negation elimination (or excluded middle) or an operator exchange the intuitionist might reject. Though see fn.11, and in any case the point I go onto make with the principle extends beyond the Kantian context.)

not an insubstantial assumption). *Then* would E-CLOS stand in tension with KPE in the Kantian context? The answer, it seems to me, is still no.

Strictly speaking what Kant denies is that the world of appearances could ever be given in experience *as* a whole (A522/B550ff.). From this it follows from Kant's conditions on knowledge only that we cannot know *of* a truth about the world of appearances as a whole *that it is such*. For the great big (though still finite) conjunction of all empirical truths to qualify as a truth Kant thinks we cannot know, then, it would have to be more than a long list concerning physical objects, their properties, the relations between them and so forth. It would also have to indicate self-referentially that it is a truth about the world of appearances as a whole – it would have to include some kind of 'and that's all' clause. It is not at all clear that such a clause could be merely another empirical truth. If not, then once again, E-CLOS would not apply and there would be no tension between Kant's epistemic restrictions on such cases and KPE.

Note, however, that this is not to say that such a clause would be an a priori truth. (This will be important in the next section.) 'And that's all' doesn't look especially a priori and plausibly it is precisely this kind of clause that Kant objects to in statements that would purport to be about the world as a whole – he thinks they are meaningless, without genuine cognitive significance, and therefore classifiable neither as empirical nor as a priori. That is, Kant's epistemic restrictions on cases like the above are a *result* of his anti-realism and do not stand in tension with it just because of the logic of certain classes of truth.

Our Fitch-Church style proof looks in good shape so far. And in fact, things are even worse than this. For it turns out that we can also prove that all *a priori* truths are knowable and that there are no unknown *a priori* truths.

1.2

Not only is empirical truth closed under conjunction, it also dominates in conjunction with a priori truth. That is, conjoining an a priori truth with an empirical truth yields an empirical truth. Where A is the a priori truth operator 'it is an a priori truth that':

(E-DOM) $A\phi \land E\psi \rightarrow E(\phi \land \psi)$

For if experience is necessary for knowledge of some truth, then surely it will also be necessary for knowledge of the conjunction of this truth and some other truth, even if knowledge of the other truth on its own would not require experience. Otherwise put, how could conjoining an empirical truth with some other truth discharge the demand for experience? (So like E-CLOS and EK, E-DOM does not rely on any problematic conflation of the metaphysical and epistemic modalities. But for reasons that will become clear below, it is worth noting that contingency also dominates in conjunction with necessity in any normal modal logic in which necessity entails truth.[16])

Now take an arbitrary a priori truth p and an arbitrary empirical truth q. By E-DOM, the conjunction of p and q is itself an empirical truth, and so by KPE knowable. But knowability, just like possibility and knowledge individually, distributes over conjunction. A conjunction is knowable only if its conjuncts are: $\Diamond K(\phi \wedge \psi) \rightarrow \Diamond K\phi \wedge \Diamond K\psi$. Thus p is knowable. But our choice of p was arbitrary – it could have been any a priori truth. We have a priori knowability:

$$(\text{KPA}) \quad A\phi \rightarrow \Diamond K\phi$$

With KPA in place, we can show that if some a priori truth is unknown, then it must be an empirical truth that that truth is unknown:

$$(\text{AK}) \quad A\phi \wedge \neg K\phi \rightarrow E\neg K\phi$$

The reasoning here exactly parallels that by which we eventually derived EK from KPE. We assume $A\phi \wedge \neg K\phi$ for conditional proof. Applying KPA to the left conjunct gives us $\Diamond K\phi$, which in turn gives us $\neg\Box\neg K\phi$. Conjoining this with the right conjunct of our initial assumption, we get $\neg K\phi \wedge \neg\Box\neg K\phi$. Now we appeal to the factivity of *a priori* truth, or $A\phi \rightarrow \phi$. This allows us to infer ϕ from the left conjunct of our initial assumption, and the rest is the same as before. Putting these results together gives us $\phi \wedge \neg K\phi \wedge \neg\Box\neg K\phi$,

---

[16] Suppose that there is some world w in which it is false that $\phi \wedge \neg\Box\phi \wedge \Box\psi \rightarrow \phi \wedge \psi \wedge \neg\Box(\phi \wedge \psi)$. Then by the truth of the antecedent there is some world v accessible from w in which $\neg\phi$, though ϕ is true in w, as is ψ if necessity entails truth. Therefore the falsity (in w) of the antecedent must be due to the falsity (in w) of $\neg\Box(\phi \wedge \psi)$, so $\phi \wedge \psi$ and in particular ϕ must be true in every world accessible from w – contradiction. (Again it is not clear whether an intuitionist would accept this proof, but again see fn.11 and again the point I go on to make with it extends beyond the Kantian context.)

which says that it is a contingent truth that some truth is unknown. From our Kripke-proof claim that all contingent truths about truths being unknown are empirical truths, we can infer $E\neg K\phi$. And discharging our initial assumption through conditional introduction yields AK.

We are now in a position to run another Fitch-Church style proof to show that no a priori truth is unknown. The first stage is exactly as before, proving the standard Fitch-Church lemma that the Moore proposition is unknowable, or $\neg\Diamond K(\mathrm{p}\wedge\neg K\mathrm{p})$. The second stage is very similar:

(7') $\exists p\ (Ap\wedge\neg Kp)$ assumption

(8') $A\mathrm{p}\wedge\neg K\mathrm{p}$ 7', existential instantiation

(9') $A\mathrm{p}\wedge E\neg K\mathrm{p}$ 8', AK to yield right conjunct

(10') $E(\mathrm{p}\wedge\neg K\mathrm{p})$ 9', E-DOM

(11') $\Diamond K(\mathrm{p}\wedge\neg K\mathrm{p})$ 10', KPE

(12') $\neg\exists p\ (Ap\wedge\neg Kp)$ lemma, 7', 11', discharging ass.

Kant is in trouble. *If* he thinks that all empirical truths are knowable, then he must concede not only that there are no unknown empirical truths but also that all a priori truths are knowable and that there are no unknown a priori truths. He can maintain neither empirical nor a priori humility.

This is an interesting result, and it generalises. Call any class of truths that contains the Moore proposition a Moore class. The wider lesson of the original Fitch-Church lemma is that knowability cannot consistently be maintained for Moore classes. The wider lesson of the last section was that for any class of truths C that obeys principles corresponding to EK and E-CLOS, if a humility claim holds for C then C is a Moore class.[17] By the lesson of the Fitch-Church lemma this means that for such classes knowability entails a failure of humility. And the wider lesson of the present section is that for any classes of truths C and D related by principles corresponding to E-DOM and AK (such that C-truths dominate in conjunction with D-truths and if a D-truth is unknown then it is a C-truth that the D-truth is unknown, rather than vice versa): knowability in C entails a failure of

---

[17] By 'corresponding' principles I mean ones in which the relevant truth operators are replaced by those defining membership of the relevant class.

humility in *D*. This is because humility in *D* would entail that *C* is a Moore class, but being a Moore class is incompatible with knowability by the Fitch-Church lemma. Call such pairs of classes Moore pairs. To restate the key point: a Moore pair is an ordered pair of classes of truths < *C, D* > such that knowability in *C* entails a failure of humility in *D*.

We can now see that the result of the last section is just the special case of this new result in which *C* and *D* are the same class. Otherwise put, principles corresponding to EK and E-CLOS are special cases of principles corresponding to AK and E-DOM in which the truth operators are the same. What we have seen, then, is that empirical truth forms a Moore pair both with itself and with a priori truth.

And it is easy to see that the same holds for contingent and necessary truth. Contingent truth forms a Moore pair both with itself and with necessary truth. For contingent truth obeys principles corresponding to EK and (in any normal modal logic) E-CLOS, and contingent truth and necessary truth are suitably related by principles corresponding to AK and (in any normal modal logic in which necessity entails truth) E-DOM. Thus one cannot (in any normal modal logic in which necessity entails truth) maintain knowability for contingent truth without also being committed to a failure of humility for both contingent truth and necessary truth.[18]

Restricted knowability principles can systematically spread, and with them spreads the associated failure of humility. And although we have not relied on any dubious conflation of the epistemic and metaphysical modalities, we have seen that there are deep, structural similarities between how each functions in a framework of knowability.

But are we being fair to Kant? Arguably not.

Let us reconsider KPE. It is already a restricted knowability principle – it says only that all empirical truths are knowable. But plausibly the passage cited in §1.1 suggests an even more restricted principle, namely that all *purely* empirical truths are knowable, where a purely empirical truth is an empirical

---

[18] It is a relevant further question whether the same holds for analytic and synthetic truth (or indeed any other significant pair). I cannot explore the issue here but it does seem plausible that synthetic truth will form a Moore pair both with itself and with analytic truth.

truth that consists solely of empirical truths. In particular, purely empirical truths have no a priori conjuncts. After all, this fits just as well with the example Kant actually gives: 'That there could be inhabitants of the moon, even though no human being has ever perceived them, must of course be admitted' (A492-3/B521). (A similar move could be made for contingent truth and the considerations that follow would be parallel.)

To avoid having to introduce a new symbol, we can affect the proposed further restriction on Kant's knowability principle by reinterpreting E as 'it is a purely empirical truth that'. What effect does this have on our results?

In effect what happened above was that the empirical domain infected the a priori domain. This was possible because of the way Kant draws his initial distinction. He clearly wants the distinction to be exhaustive so that all truths are either empirical or a priori. But this forces us to put conjunctions of empirical truths and a priori truths in the empirical camp, which in turn caused a certain leakage between the two domains. Our new class – the purely empirical – restores the division.

Specifically, our proposed reinterpretation of E invalidates E-DOM. Conjoining a purely empirical truth with an a priori truth does *not* yield a purely empirical truth, since one of the conjuncts in the resulting conjunction will be an a priori truth and purely empirical truths have no a priori conjuncts. *Purely* empirical truth therefore does *not* form a Moore pair with a priori truth, so maintaining knowability for purely empirical truth will not entail anything about knowability or humility when it comes to a priori truth.

Nor does purely empirical truth form a Moore pair with empirical truth. Conjoining a purely empirical truth with an empirical truth *could* yield a purely empirical truth. If, that is, the empirical truth in question happened to be a purely empirical truth itself. But not all empirical truths are purely empirical truths – some of them have a priori conjuncts – so conjoining a purely empirical truth with an empirical truth might not yield a purely empirical truth. The relevant permutation of E-DOM fails here too, and maintaining knowability for purely empirical truth entails nothing about knowability or humility with regard to empirical truth more generally.

Nevertheless, purely empirical truth *does* form a Moore pair with itself. That is, the proof from §1.1 still goes through under the new interpretation of E and knowability for purely empirical truth entails a failure of humility for purely empirical truth. For purely empirical truth clearly obeys E-CLOS. Conjoining one purely empirical truth with another yields a purely empirical truth – where would an a priori conjunct come from? And while EK is not quite so straightforward, it too still stands.

One of the steps in the initial attempt at deriving of EK from KPE under the old interpretation of E was to assume that all contingent truths are empirical truths. Even if Kant does hold this assumption, this move is no longer available. Whatever the relation between contingent truth and empirical truth generally, it is not plausible that all contingent truths are *purely* empirical truths. Some contingent truths will have a priori conjuncts – recall that contingency dominates in conjunction with necessity, and presumably at least some necessary truths are a priori truths.

On the basis of Kripkean doubts but also doubts about whether Kant really does conflate the metaphysical and epistemic modalities, I then weakened the operative assumption to the claim that all contingent truths *about some truth being unknown* are empirical truths. This too sufficed to derive EK from KPE under the old interpretation of E. But again, this too is suspicious under the new interpretation of E. Must all contingent truths about some truth being unknown be *purely* empirical truths? Arguably not if the unknown truth in question is anything other than itself a purely empirical truth. Fortunately, this is not due to a return of the Kripkean doubts and it is easy to build this further restriction into the assumption. That is, if the Kripkean can rest content with $\phi \wedge \neg K\phi \wedge \neg \Box \neg K\phi \rightarrow E\neg K\phi$ under the old interpretation of E, as we saw in the last section she can, then there is no reason she cannot also allow $E\phi \wedge \neg K\phi \wedge \neg \Box \neg K\phi \rightarrow E\neg K\phi$ under the new interpretation of E. If it is a contingent truth that some purely empirical truth is unknown, then it is a purely empirical truth that that purely empirical truth is unknown. And this is obviously still sufficient to allow us to derive EK from KPE. Using KPE to derive $\neg \Box \neg K\phi$ from $E\phi \wedge \neg K\phi$ in the normal way, we then directly conjoin these to yield the antecedent of our new assumption, $E\phi \wedge \neg K\phi \wedge \neg \Box \neg K\phi$, and the rest runs as before. The result is intuitive. If some purely empirical truth is unknown (but all purely

empirical truths are knowable), where would an a priori element come from in the truth that that purely empirical truth is unknown?

With EK and E-CLOS secure under the new interpretation of E as 'it is a purely empirical truth that', where a purely empirical truth is an empirical truth with no a priori parts, the proof from §1.1 stands. Now, it is not in fact clear whether Kant would allow that there could even be any such thing as a purely empirical truth. But the point here is just that it would not help in any case. We would be left with the result that if Kant thinks that all purely empirical truths are knowable, then he cannot consistently maintain that some purely empirical truths are unknown. And this is hardly more palatable than the previous results. *If* there is any such thing as a purely empirical truth, then surely there are countless unknown ones. For instance, assuming these are reasonable candidates for purely empirical truths, either it is a purely empirical truth that the number of hairs on my head as I write this is 100,001, or it is a purely empirical truth that it is not the case that the number of hairs on my head as I write this is 100,001. But whichever it is, no-one will ever know – I'm not going to bother to count them and nor is anyone else.

<div align="center">**2.1**</div>

Perhaps this is just as it should be. Perhaps Kant's position is unpalatable, even inconsistent. This would not be the first time such an accusation has been made. Or another option at this point would be to deny altogether that Kant was an anti-realist.[19] I think both of these responses are too quick, at least with regard to the present problem, and in the rest of this paper I will argue for a more measured, and to my mind more interesting, solution. Of course my claim is not that Kant recognised and pre-emptively resolved the knowability paradox as manifested in our Fitch-Church style proof. Rather it is that there is a reading of Kant on which the arguments developed in the previous sections cannot gain their initial foothold. For initial appearances to the contrary, Kant is not committed to empirical knowability. In fact he carefully, even explicitly, eschews it.

We saw at the beginning of §1.1 that Kant certainly seems to define *some* class of truth in terms of possible experience. This is one of the things that

---

[19] See Abela (2002) and Van Cleve (1999, 212-25).

makes it difficult to deny that Kant espouses *some* form of anti-realism. Note, however, that to find an overt commitment to *knowability* in particular in such contexts, it was also required that we conceive of Kantian experience as a kind of knowledge. Only then does a definition of truth in terms of possible experience look like a direct commitment to knowability. This is the standard reading of Kantian experience. It is common across otherwise very diverse interpretations.[20] And this is the source of one of the key points of interest in the results of the previous sections – a serious philosophical problem appears to arise for Kant's view regardless of whether or not we interpret his transcendental idealism in a metaphysically moderate way. What I want to suggest, then, is that the standard reading of Kantian experience is not always correct. At least it does not tell the whole story.

Kant sometimes uses the term 'experience' ('*Erfahrung*') to denote something far more elaborate than knowledge, at least as we understand it and in a sense relevant for knowability. He talks about 'the one all-encompassing experience' (e.g. at A582/B610) and appears to have more than a mere collection or unified serial expansion of particular experiences in mind:[21]

> There is only *one* experience, in which all perceptions are represented as in thoroughgoing and lawlike connection, just as there is only one space and time, in which all forms of appearance and all relation of being or non-being take place. If one speaks of different experiences, they are only so many perceptions insofar as they belong to one and the same universal experience. (A110)

Call experience in this sense Experience*. Drawing on the work of Nick Stang (2012; 2016, chapter 7; forthcoming),[22] I want to propose that

---

[20] For a small but indicative sample, see Bird (1962), Strawson (1966), Walker (1978), Guyer (1987), Van Cleve (1999), Allison (2004), and Allais (2015).

[21] See also Bxli, A146/B185; *Prolegomena* (4:292, 320); and especially *Opus Postumum* (21:85; 22:353, 457, 550, 552, 611). Plausibly the same idea, although less explicitly, is present in Kant's talk of 'experience in general' (e.g. at A225/B272) and of the 'unity of experience' (e.g. at A229-30/B282). And finally, in the *Metaphysical Foundations of Natural Science*, Kant talks of 'a *determinate* concept of experience (which *unites all appearances*)' (4:560, my emphasis).

[22] There is also much relevant material in Friedman (2013, chapter 4), in particular in his discussion of the role in Kant's theory of the idea of absolute space.

Experience* is that lawful representation of the world (of appearances) which is maximally justified by the totality of human perceptions. By 'lawful' I mean that Experience* represents the world as maximally systematic and unified in virtue of being governed by the synthetic a priori principles of the pure understanding, the synthetic a priori laws of pure natural science, and the true empirical laws of natural science. In saying that Experience* involves the *true* empirical laws, I do not mean that which putative empirical laws are the true empirical laws is somehow determined prior to the determination of the content of Experience*, with Experience* then successfully picking these out. Rather the content of Experience* *is itself* what determines which are the true empirical laws. Indeed, accordance with Experience* is *the* criterion of (appeariential) truth *in general*. For Experience* is *the* true representation of nature – it is final science.

First I will say a little in elaboration and defence of this conception and its centrality to Kant's project. Then I will explain how it is relevant in the present context.

Kant's concern in the *Critique of Pure Reason* is with the 'conditions of the possibility of experience', and in particular to show that these 'are at the same time the conditions of the possibility of the objects of experience' (A158/B197). Now on the current proposal, 'experience' here is ambiguous, but not in a problematic way. For according to Kant, if a lawful representation of the world or its justification by the totality of human perceptions were not even possible – for instance because transcendental chaos rather than transcendental affinity obtained, because cinnabar 'were now red, now black, now light, now heavy' (A100)[23] – then nor would it be possible to have so much as an empirical representation with objective purport. 'Nothing is an object *for us*', Kant says, 'unless it presupposes the sum total of all empirical reality as condition of its possibility' (A582/B610). And even more to the point a little later:[24]

---

[23] For discussion see Westphal (2005; 2006).
[24] It is worth pointing out that although they come from the Dialectic, these points are made in Kant's *positive* account of the *legitimate* uses of the ideas of pure reason, albeit in regulative rather than constitutive principles. See Friedman (1992) for a very relevant discussion of the general distinction, and Stang (2012) and Friedman (2013, 544-62) on the two cases at hand.

without [the systematic unity of nature] we would have no reason, and without that, no coherent use of the understanding, and, lacking that, no sufficient mark of empirical truth; thus in regard to the latter we simply have to presuppose the systematic unity of nature as objectively valid and necessary. (A651/B679)

For Kant, the possibility of Experience* is a necessary condition of the possibility of experience, be the latter knowledge in the everyday sense or perception as of objects in the everyday sense. Thus by the transitivity of the relation, any condition on the possibility of Experience* is also a condition on the possibility of experience. In this way we can take Kant's concern with the 'conditions of the possibility of experience', along with his claim that these 'are at the same time the conditions of the possibility of the objects of experience', as a concern with Experience*. And we can do so *without* having to argue that he has specifically Experience* in mind when he uses the word 'experience' in such contexts. The thought is then threefold.

First that Kant's articulation of the conditions of the possibility of experience in, for instance, the synthetic a priori principles of the pure understanding of the *Critique*, or indeed his specification of these principles to the objects of outer sense through the determination of the concept of matter in the *Metaphysical Foundations of Natural Science*, amounts to an articulation of the essential, abstract structure of Experience*. Second that this in turn amounts to an articulation of the essential, abstract structure of *nature itself* (considered materially as the 'sum total of all objects of experience', 'the whole of all appearances, that is, the sensible world'[25]). And third that the *particular* way in which this abstract structure is *actually* manifested – the particular way that the sensible world happens to actually be and the particular content that Experience* happens to actually have – is a function of the totality of human perceptions.

That, in a nutshell, is how Experience* is relevant to Kant's Critical project. How is it relevant in the current context? This time the thought is twofold.

First, we apply Experience* in interpreting those passages in which Kant seems to express anti-realism and in particular knowability. This is highly

---

[25] See *Prolegomena* (4:296) and *Metaphysical Foundations* (4:467) respectively.

plausible. In the paragraph immediately preceding the inhabitants-of-the-moon passage quoted in §1.1, Kant says something very similar except that he refers explicitly to *one* experience:

> In space and time, however, the empirical truth of appearances is satisfactorily secured, and sufficiently distinguished from its kinship with dreams, if both are correctly and thoroughly connected up according to empirical laws in one experience. (A492/B520-1)[26]

Likewise for another passage central to the anti-realist reading of Kant, this time from the Postulates of Empirical Thinking in General. A metaphysical rather than semantic turn of phrase is employed here. But after defining the actual as 'That which is connected with the material conditions of experience' (A218/B266), Kant goes onto clarify that this is no straightforwardly idealist *esse est percipi* claim and again relevantly qualifies the conception of experience in play:

> The postulate for cognizing the actuality of things requires perception, thus sensation of which one is conscious – not immediate perception of the object itself the existence of which is to be cognized, but still its connection with some actual perception in accordance with the analogies of experience, which exhibit all real connection in an experience in general. (A225/B272)

Indeed, there is much to suggest that Kant is talking about Experience* throughout the Postulates. He repeatedly uses phrases like 'the sum total and context of a single experience, of which each given perception is a part' and 'a single all-encompassing experience'. And in providing his definitions of the three modal categories in their empirical use, Kant denies that the sphere of possibility includes more than that of actuality (A230-3/B282-4). This suggests that he is working with a narrower-than-nomological conception of modality on which not only the deterministic exceptionless laws are held fixed but also certain states. That is, Kant is explaining how the modal

---

[26] See also B279: 'whether this or that putative experience is not mere imagination must be ascertained according to its particular determinations and through its coherence with the criteria of all actual experience'.

categories function in Experience*. For it is only once certain states are held fixed as well as the deterministic exceptionless laws that there is nothing possible that is not also actual, and it is precisely Experience* that fixes these states.

The second part of the thought is then as follows. Experience* is a mere *ideal*. Its content is that of the final scientific theory of the world of appearances. Through continued empirical inquiry, conducted in accord with methods prescribed by other regulative ideals of pure reason, we can gradually and with ever increasing accuracy approximate to such a theory. But we can never attain our goal. For we can never be given the totality of human perceptions. From our finite standpoint, there could always be more evidence, always more experience (little 'e', no star). There is, therefore, no knowable fact of the matter as to which theory is maximally justified. And since, for Kant, a final science would contain within itself assurance of its own finality, such a thing lies in principle beyond our epistemic reach.[27]

One might object at this point that an ideal, in this sense, is precisely *not* possible. So if Experience* is an ideal, how can one of Kant's central concerns be with the conditions for its possibility? More to the point, I said above that the possibility of Experience* is a necessary condition of the possibility of experience. But if Experience* is an ideal and therefore not possible, it then follows that experience is not possible.

This objection conflates metaphysical impossibility with the epistemic sense in which Experience* is an in principle unachievable ideal. Experience* is metaphysically possible. Indeed it is metaphysically actual. Experience* is the complete, accurate representation of nature and nature is actual. Experience* remains, however, in principle beyond our epistemic reach. For we cannot possibly know what particular content it has. We can, to be sure, know its abstract structure. This is what is articulated by Kant's various a priori principles. But this only determines a range of metaphysically possible instantiations, a range of metaphysically possible Experiences*. Which particular metaphysically possible Experience* is as a matter of fact metaphysically actual remains essentially unknowable to us. To adapt a

---

[27] See, e.g., A642ff./B670ff., especially A662-8/B690-6; A323/B380; A482ff./B510ff.; A567-8/B595-6.

famous formula of Kant's, Experience* is metaphysically real but epistemically an ideal.

Now, if *this* is the sort of thing in terms of which Kant defines truth, then there is no knowability principle here. For there to be inhabitants of the moon, for example, just is for final science to entail that there are inhabitants of the moon, which is in turn for the totality of human perceptions to maximally justify a representation of the world (of appearances) according to which such inhabitants are part of it. But final science is not itself an object of possible knowledge. Thus the view in no way entails that all truths are knowable. In particular, it entails neither that all empirical truths are knowable, nor that all purely empirical truths are knowable. Indeed, since Experience* contains much particular empirical content alongside the kinds of structural a priori content mentioned above, and so would itself presumably qualify as an empirical truth (albeit not a purely empirical truth), the view is positively incompatible with empirical knowability. Does it still qualify as a form of anti-realism? The answer to this question is not yet clear and I return to it in the next section. In a strictly limited sense the view is one on which truth not only potentially but essentially transcends our cognitive capacities; in another sense, however, truth retains an essential connection to these capacities. In any case, the view is certainly a form of idealism. Truth, it says, is a function of human perception.

There is of course much more that could be said about my proposal, both in its defence and in elaboration of its details and consequences. On the face of it the view I am ascribing to Kant looks highly reminiscent of C. S. Peirce's (1934, 565) famous definition of truth as 'that concordance of an abstract statement with the ideal limit towards which endless investigation would tend to bring scientific belief'.[28] And fittingly one immediate result would be that Kant's conception of at least one kind of knowledge is fallibilist. If Experience* is the criterion of truth yet epistemically inaccessible to us with regard to its particular content, then for no belief that p, where p concerns that content, can we rule out the possibility of ¬p. Any putative instance of

---

[28] Though note that, if we understand ideal in a Kantian way here, this definition is not especially pragmatic and says something very different to Peirce's other oft-cited definition of truth as 'The opinion which is fated to be ultimately agreed to by all who investigate' (1934, 407).

knowledge of such propositions is just that – putative. The view entails that at least this degree of anti-Cartesianism is present in Kant's epistemology. But I must leave a fuller account for another occasion. Here I want to focus on a remaining issue that relates specifically to Fitch-Church style reasoning.

If what I have said is right, then there is no longer reason to attribute to Kant any of the knowability principles we have explored. But arguably it remains plausible to attribute to Kant the claim that all purely empirical truths can be the objects of justified belief, and this principle can appear just as problematic.[29]

2.2

So far I have focused on experience. But in the passages we have seen, Kant also connects truth, via reality and actuality, to perception. On the proposed reading, this is entirely natural – there is an intimate connection between Experience* and perception. Thus in the moon passage quoted in §1.1, Kant says 'to call an appearance a real thing prior to perception means either that in the continuation of experience we must encounter such a perception, or it has no meaning at all' (A493/B521), and in the passage from the Postulates quoted in §2.1, 'cognizing the actuality of things requires [a suitable connection to] perception' (A225/B272). The problem is that, unlike Experience*, perception is no mere ideal.

All can agree that Kantian perception is not knowledge, so the worry here is not that we might have a return of the knowability principle. But because perception is no mere ideal, the view does appear to entail a more straightforward form of anti-realism than what we had at the end of the previous section. For suppose, as seems plausible, that perception always yields evidence, even if not conclusive evidence or knowledge. Then in connecting the very meaning of reality and actuality to possible perception, Kant seems to be saying that the notion of a purely empirical truth that is

---

[29] In fact there remains one further issue related to Fitch-Church style reasoning. For all I have said here, it is not clear whether Kant might still be committed to *a priori* knowability, and if he is, whether this entails a failure of a priori humility. The proof from §1.2 relied on empirical knowability and the kind of proof we saw in §1.1 would not go through for a priori truth. See AK – propositions of the form $\phi \wedge \neg K\phi$ are not a priori truths even when ϕ is an unknown a priori truth. But there is an alternative route that goes via a KK-principle instead. I discuss a priori knowability on its own terms in Stephenson (ms.)

evidence-transcendent makes no sense.

And we might go further. Arguably, if we can perceive a state of affairs to be the case, then we can at least be justified in believing that state of affairs to obtain.[30] Of course such a belief might turn out to be false – perception here can be non-veridical.[31] Or we might perceive something and not believe our eyes, or there might be epistemic defeaters in the area that undercut our justification. And even if we do form a true justified belief on the basis of our perception, there's always Gettier. But perception *can* confer at least defeasible warrant. So it looks like Kant is committing to the claim that all purely empirical truths can be the objects of *justified belief*. Where *JB* is a justified belief operator for 'someone at some time has a justified belief that':

(JBE) $E\phi \rightarrow \Diamond JB\phi$

Our question is: what effect would replacing KPE with JBE have on the proof from §1.1?

It would not significantly affect the second stage of the proof. The revised conclusion, that there is no purely empirical truth in which no one ever has a justified belief is just as bad; the closure principle for purely empirical truth is not affected at all; and given that all purely empirical truths must permit of justified belief, it remains eminently plausible that it will be a purely empirical truth when some purely empirical truth happens to be one in which no one ever has a justified belief. This time the natural place to look is the first stage of the proof. It was assumed in this stage that knowledge is factive, and this looks suspicious for justified belief, especially when we have arrived at the justified believability condition through a perceivability condition according to which perception can be non-veridical.

Denying factivity is indeed crucial for avoiding a justified belief version of the Fitch-Church style proof, but it is not alone sufficient. If we hold the other rules employed in the first stage fixed – distribution over conjunction, necessitation, etc. – then the first stage of the proof can still be run if we just replace factivity with the following reflection principle:

---

[30] For a recent discussion, see Siegel and Silins (forthcoming).
[31] For discussion of Kant's theory of non-veridical perception and what it tells us about his model of the mind more generally, see Stephenson (2011; 2015).

(RP) If someone at some time justifiably believes that no one ever justifiably believes that ϕ, then at that time she does not herself justifiably believe that ϕ.

On the face of it, RP might look quite attractive. For suppose that someone has a justified belief that p and is aware of this. Then it might take minimal reflection on her part to realise that it is not the case that no one ever has a justified belief that p. Or alternatively, if we assume that if one justifiably believes that no one ever justifiably believes that ϕ, then one also justifiably believes that one does not oneself justifiably believe that ϕ – a kind of qualified closure under entailment – we could weaken our replacement reflection principle RP to the following, which might be thought of as a kind of qualified, subjective factivity:

(RP') If someone at some time justifiably believes that she does not herself at that time justifiably belief that ϕ, then in fact at that time she does not herself justifiably belief that ϕ.

We need not formalize any of this.[32] The justified belief version of the original Fitch-Church lemma says that it is impossible for a subject at a time to justifiably belief both that p and that no one ever justifiably believes that p. If we are to avoid the proof, what we need is a conception of justified belief according to which this is not impossible. As we shall see, such a conception would also produce counterexamples to the above reflection principles.

Consider the following case, due to Tim Button (2013, 102-3):

Kate has obtained a belief that [p], via an extremely reliable method. However, Kate does not believe that this method is extremely reliable. She thinks (mistakenly) that it is highly prone to mistakes, although she thinks that it is better than nothing. So Kate thinks (mistakenly) that her belief that [p] falls short of *justification*. Thus Kate has a justified true belief that [p], though she (falsely) thinks that her belief is unjustified.

   In fact, Kate's caution is very sensible. All the evidence available to Kate suggests that she should *not* place much faith

[32] For formalization, see Kelp and Pritchard (2009).

in the method by which she came to believe that [p]. Now, Kate
might also be aware that there are *other* methods for
determining whether or not [p]. However, Kate is basically
certain that no one will ever implement them, and rightly so:
these other methods are extraordinarily arduous. So Kate comes
to a justified (but false) belief that nobody will ever have a
justified belief that [p].

There are two crucial features of the conception of justification at work in
this case. First, that what makes Kate's true belief that p a justified true belief
is that the method via which she obtained her belief was an extremely
reliable one. It makes no difference that Kate is not in a position to know
that her method was a good one, nor indeed that she believes otherwise.
That is, the conception of justification at work here is at heart an externalist
one. Yet, and this is the second point, it is also non-factive. For Kate also has
a justified false belief, namely that nobody will ever have a justified belief
that p. To be sure, if this false belief is to be justified, then her evidence that
her own method of coming to believe that p is highly prone to mistakes and
that other methods would be prohibitively arduous must in fact be good
evidence. But this is not incompatible with it being ultimately misleading.

What I now want to suggest, drawing this time on the work of Andrew
Chignell (2007),[33] is that this conception of justification fits precisely Kant's
own.

Kant talks a lot about grounds for assent.[34] Typically he has in mind things
like perception, memory, testimony, certain assent-types themselves, such as
belief, and also inference and argument. Such grounds are 'subjective' when
they motivate a subject to adopt some attitude of assent. Thus the method
by which Kate came to her belief that p would qualify as a subjective ground
for assent, as would whatever combination of reasoning and evidence led her
to believe that nobody will ever have a justified belief that p. What we
would call justification comes into the picture when subjective grounds for
assent are also what Kant calls 'objectively sufficient'. So what does it take,
according to Kant, for a ground to be objectively sufficient, for it to confer

---

[33] See also Westphal (2004).
[34] See especially A820-31/B848-59, *Jäsche* (9:66-73), *Blomberg* (24:142-9, 194-7).

justification?

The texts are diffuse and obscure but Chignell (2007, 39-44) argues forcefully that a ground for assenting to some judgement is objectively sufficient just in case it renders that judgement objectively probable to a moderate-to-high degree.

As Chignell fully admits, further details are difficult to determine, in particular the precise notion of objective probability in play here. But it is at least clear that probability (*Wahrscheinlichkeit*) is the correct measure, and moreover that Kant has is mind *some* objective rather than subjective kind of probability. He links the notion to truth, objective validity, and the object, and is careful to distinguish it from mere plausibility (*Scheinbarkeit*), which he links to the subject and subjective validity. For example, student lecture notes have him saying 'Probability and improbability are objective grounds. Plausibility, however, is a subjective ground; that is, probability and improbability lie in the object itself, in the thing that is to be cognized' (24:145). And 'plausibility can alter greatly, for its certainty is grounded in the subject, but probability always remains unchanged, and this merely because it is grounded in the object' (24:195).

What is crucial for our purposes is that this is enough to show that Kant's conception of justification is externalist in the relevant way. There are facts about what particular relations of probabilification hold between particular grounds and particular judgements. These in turn determine facts about whether a subject's grounds for some assent are objectively sufficient, about whether the grounds in question confer justification on the assent in question. However, all of this holds 'whether I have insight into these grounds or not' (24:194). There is no reason to demand and nor would one expect that all or even most of these facts are always or even often available to the subject.[35]

---

[35] I should note that Chignell also argues that there remains an emphatic nod to the internalist in Kant's model – subjects must be able to cite their objectively sufficient grounds, albeit not *that* their grounds are objectively sufficient. This is not required by Button's model, but it is compatible with it. Kate is both motivated to believe that p by her method and aware that she was so motivated, even though she falsely yet justifiably believes that her method was not reliable enough to confer justification. That is, Kate *is* in a position to cite the grounds for her belief that p, and since her grounds are good grounds, she is

Matters are similar for the non-factivity of Kant's conception of justification. It is admittedly a little unclear whether Kant really thinks that a moderate-to-high degree of probability is required, rather than merely anything above .5. Sometimes what he says is ambiguous, sometimes what he says unambiguously picks out the latter, and ultimately Chignell argues in favor of the former as much on grounds of charity and consistency with what Kant says elsewhere about precise probability measures. But again it is at least clear that the degree of probability conferred on a judgement by a ground need not be 1 in order for that ground to qualify as objectively sufficient for assent to that judgement. To be sure, the gold standard of an objectively sufficient ground for some assent is a demonstrative proof that confers probability of 1 on the judgement assented to, but such is by no means necessary and in most contexts not even possible. Kant's example is historical as opposed to mathematical judgements (24:733). We can have objectively sufficient grounds to assent to both kinds of judgement. Thus we can have justified beliefs in both kinds of judgement. But only in the latter case could grounds confer a probability of 1 on the content of our assents.

Kant's conception of justification is one on which both of the above reflection principles – RP and RP' – come out false. They presume too much transparency regarding a subject's access to the justificatory status of her own beliefs. Consider, for instance, what made RP prima facie attractive. That if a subject has a justified belief that p and is aware of this, then a little reflection would lead her to realize that it is not the case that no one ever has a justified belief that p. But on Kant's model, there is no guarantee that a subject *can* always become aware of the fact that she has a justified belief, for this will likely depend on facts beyond her ken. More specifically, Kant's conception of justification is one on which it is possible for a subject at a time, such as Kate, to have a justified belief both that p and that no one will ever have a justified belief that p. For Kant, the justified belief version of the Fitch-Church lemma is false, and justified believability principles entail nothing about humility. At least if we assess him on his own terms, Kant can consistently maintain both that all purely empirical truths can be the objects of justified belief and that there are some purely empirical truths in which no one ever has a justified belief.

thereby in a position to cite good grounds. Kant's model, if Chignell is right, would just strengthen what in Button's model is matter of fact into one of necessity.

Conclusion

In §1.1 I argued that the standard reading of Kantian experience as a form of knowledge has Kant committed to the in principle knowability of all empirical truths, and that this entails that there are no unknown empirical truths by a modified version of the familiar Fitch-Church proof. In §1.2 I extended the result to a priori truths and drew some general philosophical lessons about the way certain classes of truth interact in a framework of knowability. Ultimately this extension might not be entirely fair to Kant, but even without it he is left committed to an epistemic optimism with which he would, as a self-professed empirical realist, be very uncomfortable. In §2.1 I proposed a modification of the standard reading according to which Kant sometimes means by 'experience' not a form of knowledge at all but rather a highly technical conception of a final science, the full and precise content of which beings like us cannot even in principle come to know. On this reading, Kant is not committed to any form of empirical knowability and the proofs from §1 no longer go through. Arguably, Kant would remain committed to a principle of justified believability for purely empirical truths, but I argued in §2.2 that he has just the resources in his externalist, non-factive conception of justification to avoid Fitch-Church style reasoning when it comes to this weakened principle.

Distinguishing experience as a form of knowledge from experience as an epistemic ideal science yields a view with both realist and idealist characteristics that is at heart an interesting and complex form of anti-realism. There is much more work to be done in exploring and defending my proposal, in disentangling these different notions of experience and mapping how they relate and what role each plays in Kant's theory. I have focused here on articulating a serious philosophical problem that afflicts Kant under a broad range of other interpretations, and on showing that my proposal allows him to avoid it. So far the prospects are good.[36]

References

Abela, P. (2002). *Kant's Empirical Realism*, Oxford: Clarendon Press.

Allison, H. (2004) *Kant's Transcendental Idealism* (2nd revised and enlarged edn.), London: Yale University Press.

Allais, L. (2003). 'Kant's transcendental idealism and contemporary anti-realism', *International Journal of Philosophical Studies*, 11: 369-392.

- (2015). *Manifest Reality: Kant's Idealism and his Realism*, Oxford: Oxford University Press.

Ameriks, K. (1982). *Kant's Theory of Mind*, Oxford: Clarendon Press.

Baldwin, T. (2001). 'Kantian Modality', *Proceedings of the Aristotelian Society*, suppl. 76: 1-24.

Bird, G. (1962). *Kant's Theory of Knowledge*, London: Routledge.

Brouwer, L. E. J. (1975). *Collected Works I* (Ed. A. Heyting), Amsterdam: North-Holland.

Button, T. (2013). *The Limits of Realism*, Oxford: Oxford University Press.

Chignell, A. (2007). 'Kant's Concepts of Justification', *Noûs*, 41 (1): 33–63.

Divers, J. (1999). 'Kant's Critique of the A Priori', *Pacific Philosophical Quarterly*, 80:17-45.

Edgington, D. (1985). 'The Paradox of Knowability', *Mind*, 94: 557–568.

Friedman, M. (1992). 'Regulative and Constitutive', *Southern Journal of Philosophy*, 30: 73-102.

- (2013). *Kant's Construction of Nature*, Cambridge: Cambridge University Press.

Guyer, P. (1987). *Kant and the Claims of Knowledge*, Cambridge: Cambridge University Press.

Hanna, R. (1993). 'The trouble with truth in Kant's theory of meaning', *History of Philosophy Quarterly*, 10: 1-20.

- (2006). *Kant, Science, and Human Nature*. Oxford: Oxford University Press.

Kant, I. (1992). *Lectures on Logic.* (M. J. Young, Ed. & Trans.) Cambridge: Cambridge University Press.

- (1993). *Opus Postumum.* (E. Förster, Ed., E. Förster, M. Rosen, Trans.) Cambridge: Cambridge University Press.

- (1997). *Critique of Pure Reason.* (P. Guyer, A. W. Wood, Eds. & Trans.) Cambridge: Cambridge University Press.

- (2001). *Lectures on Metaphysics.* (K. Ameriks, S. Naragon, Eds. & Trans.) Cambridge: Cambridge University Press.

- (2002). *Theoretical Philosophy After 1781*. (H. Allison, P. Heath, Eds., G. Hatfield, M. Friedman, H. Allison, P. Heath, Trans.) Cambridge: Cambridge University Press.

Kelp, C. and Pritchard, D. (2009). 'Two Deflationary Approaches to Fitch-style Reasoning', in *New Essays on the Knowability Paradox* (Ed. Salerno, J.), 324-338.

Kripke, S. (1980). *Naming and Necessity*, Cambridge: Harvard University Press.

Mackie, J. L. (1980). 'Truth and Knowability', *Analysis*, 40: 90–92.

Moore, A. W. (2012). *The Evolution of Modern Metaphysics: Making Sense of Things*, Cambridge: Cambridge University Press.

Moran, D. (2000). 'Hilary Putnam and Immanuel Kant: Two 'Internal Realists'?, *Synthese* 123: 65-104.

Peirce, C. S. (1934). *Collected Papers of Charles Sanders Peirce*, Vol. 5. (Eds. C. Hartshorne & P. Weiss), Cambridge: Harvard University Press.

Posy, C. (1981). 'The Language of Appearances and Things in Themselves', *Synthese* 47: 313–52.

- (1983). 'Dancing to the Antinomy: A Proposal for Transcendental Idealism', *American Philosophical Quarterly* 20: 81–94.

- (1984). 'Kant's Mathematical Realism', *The Monist* 67: 115–134.

Putnam, H. (1981). *Reason, Truth and History*, Cambridge: Cambridge University Press.

- (1987). *The Many Faces of Realism*, LaSalle: Open Court.

Salerno, J. (2009). *New Essays on the Knowability Paradox* (Ed.), Oxford: Oxford University Press.

Siegel, S. and Silins N. (forthcoming). 'The Epistemology of Perception', in *The Oxford Handbook of Philosophy of Perception* (Ed. M. Matthen).

Stang, N. (2011). 'Did Kant Conflate the Necessary and the A Priori?', *Nous* 45 (3): 443-71.

- (2012). 'Kant on Complete Determination and Infinite Judgement', *British Journal for the History of Philosophy* 20: 1117-1139.

- (2016). *Kant's Modal Metaphysics*. Oxford: Oxford University Press.

- (forthcoming). 'Kant's Transcendental Idealism', *Stanford Encyclopedia of Philosophy*.

Stephenson, A. (2011). 'Kant's on Non-Veridical Experience', *Kant Yearbook*, 3: 1-22.

- (2015). 'Kant on the Object-Dependence of Intuition and Hallucination', *The Philosophical Quarterly*.

- (ms.). 'Kant, A Priori Knowability, and Tacit Knowledge'.

Stevenson, L. (1983). 'Empirical Realism and Transcendental Anti-Realism', *Proceedings of the Aristotelian Society*, suppl. 57: 131–53.

Strawson, P. F. (1966). *The Bounds of Sense: an essay on Kant's Critique of Pure Reason.* London: Routledge

Tennant, N. (1997). *The Taming of the True*, Oxford: Oxford University Press.

Van Cleve, J. (1999). *Problems from Kant*, Oxford: Oxford University Press.

Walker, R. C. S. (1978). *Kant*, London: Routledge.

- (1983). 'Empirical Realism and Transcendental Anti-Realism', *Proceedings of the Aristotelian Society*, suppl. 57: 154-77.

- (1995). 'Verificationism, Anti-Realism, and Idealism', *European Journal of Philosophy* 3 (3): 257–72.

Westphal, K. R. (2004) *Kant's Transcendental Proof of Realism*, Cambridge: Cambridge University Press.

- (2005). 'Kant, Wittgenstein, and Transcendental Chaos', *Philosophical Investigations* 28 (4): 303-23.

- (2006). 'How does Kant prove that we perceive, and not merely imagine, physical objects?', *The Review of Metaphysics* 59 (4): 780-806.

Williamson, T. (1982). 'Intuitionism Disproved?', *Analysis*, 42: 203–207.

- (1992). 'On Intuitionistic Modal Epistemic Logic', *Journal of Philosophical Logic*, 21: 63–89.

Wright, C. (2000). 'Truth as Sort of Epistemic: Putnam's Peregrinations', *Journal of Philosophy*, 97: 335–364.