

TRANSCENDENTAL KNOWABILITY AND A PRIORI LUMINOSITY

Andrew Stephenson
University of Southampton

Forthcoming in *History of Philosophy & Logical Analysis*

Abstract: This paper draws out and connects two neglected issues in Kant's conception of a priori knowledge. Both concern topics that have been central to contemporary epistemology and to formal epistemology in particular: knowability and luminosity. Does Kant commit to some form of knowability principle according to which certain necessary truths are in principle knowable to beings like us? Does Kant commit to some form of luminosity principle according to which, if a subject knows a priori, then they can know that they know a priori? I defend affirmative answers to both of these questions. And by considering the special kind of modality involved in Kant's conceptions of possible experience and the essential completability of metaphysics, I argue that the combination of knowability and luminosity principles leads Kant into difficulty.

1. Knowability, semantic anti-realism, and Kant

Suppose that all truths are knowable but that there is some truth that is unknown. Then some statement of the form $\phi \wedge \neg K\phi$ is both true and knowable. But statements of this form are unknowable. For knowledge distributes over conjunction and is factive, so any statement of the form $K(\phi \wedge \neg K\phi)$ entails some statement of the form $K\phi \wedge \neg K\phi$. Thus the claim that all truths are knowable is inconsistent with the claim that there is some truth that is unknown. If all truths are knowable then there are no unknown truths.

Call this argument the old knowability proof.¹ It is significant because it can seem as though semantic anti-realism entails a knowability principle to the effect that all truths are knowable. For suppose that meaning is essentially connected to our cognitive capacities such that the truth-value of any meaningful statement must in principle be discoverable to us. Then it can seem that for any truth ϕ we could in principle discover that ϕ is true and so in principle come to know that ϕ .² If so, and if the old knowability proof is valid, then semantic anti-realism is in trouble, for it seems absurd to think that there are no unknown truths. The old knowability proof threatens to provide a reductio of semantic anti-realism.

Of course the right kind of idealist-minded anti-realist might happily embrace the result of the proof—if they think that knowledge is suitably constitutive of truth. But what about the idealist-minded anti-realist who wants to hold on to some form of realism—if they think the world is in some robust way independent of our particular acts of knowing about it?

There are many ways to resist the above chain of reasoning. Restriction strategies deny that anti-realists need commit to the knowability of *all* truths. In general, our coming to know about the world might affect it. In particular, there seem to be statements such that performing the procedures necessary to discover their truth-value would render them false. Thus even if the truth-value of any meaningful statement must in principle be discoverable, it does not follow that all truths can in principle be discovered *to be true*, nor therefore does it follow that all truths can in principle be known. The reasoning that took us from anti-realism to a knowability principle might still go through for a restricted class of truths. But if statements of the form $\phi \wedge \neg K\phi$ behave in the pathological way just described so that the anti-realist need not commit to their knowability, then the old knowability proof no longer gains a foothold and the reductio of anti-realism does not go through.³

This is by no means the end of the matter but we already have enough to draw a useful parallel to Kant. Kant does not express a theory of meaning in terms

¹ Due to Alonzo Church and Frederick Fitch. See Salerno (2009) for an overview.

² See especially Dummett (1959).

³ See Melia (1991).

of discoverability (or verifiability, provability etc.). Nor does he restrict knowability because of pathological truths (so defined). But it is often thought that transcendental idealism is a form of anti-realism or that Kant was an anti-realist.⁴ And Kant certainly seems to restrict knowability in some way. He denies that we can have knowledge of things as they are in themselves, for instance. It is therefore natural to ask whether Kant commits to any restricted form of knowability, and if so, whether this still leads him into difficulty.

The restriction I consider here is to the synthetic a priori truths of transcendental philosophy and scientific theoretical metaphysics.⁵ These do not include empirical truths, mathematical truths, analytic truths, or claims that only play a positive role in Kant's practical philosophy. Nor do they include the claims of dogmatic, unscientific metaphysics, such as that the soul is a metaphysically simple substance. They do include truths like the principles of nature in general from the *Critique of Pure Reason*, such as the three analogies of experience, for instance that 'All alterations occur in accordance with the law of the connection of cause and effect' (B232).⁶ They also include the principles of corporeal nature from the *Metaphysical Foundations of Natural Science*, such as the three laws of mechanics, for instance that 'In all changes of corporeal nature the total quantity of matter remains the same, neither increased nor diminished' (4:541). It is not always clear just how far Kant takes the domain of such truths to extend, but it will suffice for present purposes to work with these paradigmatic examples in mind—what matters here is that Kant thinks there are truths which satisfy the conditions I go on to specify, in particular knowability. To have a label,

⁴ See e.g. Strawson (1966, 16), Putnam (1981, 6off.; 1987, 41ff.), Posy (1983), Walker (1983; 1995), Stevenson (1983), Baldwin (2001), Moore (2012, 362-6), Allais (2015, 207-31), and Stephenson (2015; 2018).

⁵ I have considered other restrictions elsewhere: to empirical truths in Stephenson (2015) and to non-epistemic truths—truths that make no reference to the kind of cognitive capacities in terms of which semantic anti-realism offers its epistemic characterizations of truth—in Stephenson (2018). In both cases I argue that Kant's position is indeed anti-realist but that it is robust in the face of knowability proofs, old or new.

⁶ References to the *Critique of Pure Reason* take the standard A/B format. References to other works are to volume and page number of the Academy edition (Kant 1900-) and are accompanied by a short English title or description. Translations are from the *Cambridge Edition of the Works of Immanuel Kant*, with occasional slight modifications. Details of the volumes consulted are at the end.

call members of the target class of truths *transcendental* truths (A146/B185, A222/B269). In one way or another, they concern necessary conditions for the possibility of experience.

Does Kant think that all transcendental truths are knowable? I will argue that he does and that a new knowability proof can be constructed even on this restricted basis. The old knowability proof won't go through because it turns on the unknowability of statements of the form $\phi \wedge \neg K\phi$, and there is no reason to think any transcendental truth takes this form. Thus a knowability principle to the effect that all transcendental truths are knowable is compatible with the unknowability, yet truth, of such conjunctions. Even if all transcendental truths are knowable, it does not follow that there are no unknown truths. Nevertheless, by appealing to some plausible additional assumptions, a new knowability proof can be constructed for the claim that there are no unknown transcendental truths. This is still a deeply problematic result. It would not be accounted for by Kant taking himself to know all transcendental truths, for the result holds for arbitrary knowers at arbitrary times, so long as certain very general conditions are satisfied.

Constructing our new knowability proof will involve drawing out two further neglected aspects in Kant's conception of a priori knowledge. First concerning the kind of modality at play in his conceptions of possible experience and the completability of metaphysics. Second concerning the luminosity of a priori knowledge.

In §2 I articulate the special conception of possibility active in anti-realist knowability principles and I compare and contrast it with a standard conception of metaphysical possibility. I also provide evidence that the conception is suitably Kantian. In §3 I argue that Kant's views about the essential completability of metaphysics commit him to a restricted knowability principle concerning transcendental truths. In §4 I suggest that Kant's conception of a priori knowledge and its relation to certainty makes it plausible to attribute to him a kind of a priori luminosity principle to the effect that, whenever a subject has a priori knowledge, then they know that they have a priori knowledge. In §5 I present our new knowability proof and explain its scope and significance.

In the remainder of the paper I consider whether this amounts to a reductio of Kant's position. In §6 I consider whether Kant might simply embrace the result of our new knowability proof by taking all knowers to always already have *tacit* knowledge of all transcendental truths. I think this is right and contains an important insight about the nature of Kant's Critical project. But it doesn't help here, for our proof can easily be adapted to show that Kant is committed to the claim that there are no transcendental truths that are not *explicitly* known by some arbitrary knower at some arbitrary time under certain very general conditions. Since this is implausible, we do indeed have a knowability proof that threatens a reductio of Kant's position. In §7 I respond to an objection to the factivity and closure principles involved in the new proof and in §8 I return to the luminosity of a priori knowledge. Kant thinks that metaphysics is our cognitive home in a way that motivates both knowability and luminosity principles. This partly constitutes his Critical-era idealist revolution in metaphysics, but it turns out to be incompatible with even the minimal realist thought that truths concerning the fundamental structure of the natural world transcend in some way our actual, explicit knowledge of them.

2. Transcendental knowability

In this section I introduce the kind of possibility that is involved in anti-realist knowability principles and formalize some axioms that govern how it functions. I also provide evidence that the conception is suitably Kantian.

The kind of possibility involved in anti-realist knowability principles is best thought of in situation-theoretic terms, rather than in terms of complete possible worlds. Relatedly, the knowability principle I will attribute to Kant is strictly stronger than the claim that for any transcendental truth there is a metaphysically possible world in which it is known. Before introducing our candidate Kantian knowability principle, then, we need to know something about the general kind of possibility that such principles involve. Call it *transcendental* possibility. As with transcendental truths, it will turn out that there is a connection to Kant's use of this term.

$\diamond K$ and $\diamond \overset{a}{K}$ are our transcendental possibility operators. $\overset{a}{K}$ is just like the familiar epistemic operator, K , except that it is restricted to a priori knowledge. \diamond , on the other hand, is quite unlike the familiar possibility operator. It is syncategorematic and does not function independently of one of the knowledge operators. Syntactically, $\diamond K$ and $\diamond \overset{a}{K}$ each function like single, unified operators. So what do they mean? Neil Tennant (2000, 829) captures the core idea:

the possibility alluded to is that of our attaining knowledge that ϕ , where ϕ already holds... it is a possibility for *us*, as knowers situated in the current state of information – or at least a possibility for some finite extension of ourselves.

As does Crispin Wright (2001, 60), in more Dummettian terminology:⁷

the range of what is feasible for us to know goes no further than what is actually the case: we are talking about those propositions whose actual truth could be recognised by the implementation of some humanly feasible process.

In effect, $\diamond K$ and $\diamond \overset{a}{K}$ are modal operators that range over states of information where the accessibility relation between states is restricted by closeness conditions derived in some suitable way from the nature of our cognitive capacities. *Our* cognitive capacities because we are not here concerned with divine knowers and the quantification over subjects implicit in the epistemic operators (and the transcendental possibility operators that involve them) is to be understood throughout as restricted to finite, discursive beings like ourselves.⁸ Since these closeness conditions, and the operators in which they figure, have an essentially epistemic dimension, I will also refer to transcendental possibility as transcendental *knowability*.

When the context of evaluation is the current state of information—how things are for us now in the actual world—we can think of $\diamond K\phi$ as saying, roughly, that given how things are with us now, in the current state of

⁷ See especially Dummett (1993, 45-6).

⁸ Our operators implicitly quantify over subjects and times. K , for instance, says that ‘it is known by someone at some time that’. The quantification over subjects should be understood throughout as restricted to adult human subjects (or at least to those with intellectual and sensible forms identical to our own).

information, it would be feasible for someone to perform investigative procedures so as to come to know that ϕ . Mutatis mutandis for $\diamond K^a \phi$. In the latter, a priori case, such procedures might involve logical deduction and mathematical reasoning (as well as, perhaps, more specifically Kantian forms of a priori investigation, such as transcendental reflection and transcendental argument). In the former, general case, which also includes the empirical case, such procedures might also include appealing to sense experience, looking around at the world, and so forth.

This gloss on transcendental knowability is rough because what qualifies as *feasible* is rather vague. Certainly it is to include what is, for us, physically and psychologically possible. But these notions are themselves vague. And we are also being asked to envisage ‘finite extensions’ of ourselves, which presumably allows some scope for development and refinement in our investigative capacities (albeit, in the current context, constrained by the essential natures of our cognitive capacities). Examining the full details of these operators goes beyond the scope of this paper.⁹ The intuitive idea will be clear enough for us to be able to recognize certain of their general features.

For instance, actual knowledge ($K\phi$) entails transcendental knowability ($\diamond K\phi$) just as actual knowledge entails metaphysical knowability ($\diamond K\phi$, where the empty diamond is the familiar metaphysical possibility operator). If it is actually known that ϕ then it must be feasible for someone to come to know that ϕ , since someone has in fact come to know that ϕ . And of course the converses do not hold. Transcendental knowability, like metaphysical knowability, is supposed to be consistent with a lack of actual knowledge. It is transcendently possible, just as it is metaphysically possible, for someone to come to know the number of hairs on my head. It would, in the relevant sense, be feasible for someone to count them. But no one will ever bother. Mutatis mutandis for a priori cases.

However, unlike metaphysical knowability, transcendental knowability entails truth. This will be crucial. $\diamond K$ and $\diamond K^a$ inherit the factivity of K and K^a . Just as ϕ must be a truth if it is known (in whatever way), so too must ϕ

⁹ See Tennant (2000; 2002) and Williamson (1992).

be a truth if it is to be feasible for someone to come to know ϕ (in whatever way). We can think of this as a quasi-stipulative constraint on what is meant by ‘feasibly knowable’ in the anti-realist context. As Wright put it above, ‘the range of what is feasible for us to know goes no further than what is actually the case’; as Tennant put it, ‘the possibility alluded to is that of our attaining knowledge that ϕ , where ϕ already holds’. Expressing the general case as an axiom schema, since it will be invoked in our new knowability proof:

$$\text{(FACT)} \quad \diamond K\phi \rightarrow \phi$$

The transcendental possibility operators are being used to capture anti-realist knowability principles. Recall from §1 that the base motivation for the anti-realist’s commitment to knowability is that *the* truth-value of any meaningful statement must in principle be discoverable *to us*. If truth-value is indexed to worlds, the question arises: *which* truth-value? The anti-realist is interested in the truth-value of statements in the actual world. The actual world, after all, is where we are. The idea is that the anti-realist is concerned to place an epistemic condition on meaning and thereby truth such that what is true is at least in part a matter of what we might actually be able to go about discovering (proving, verifying, etc.). They are no more interested in whether some contingent falsehood is known in some world quite unlike our own than they are when the knower in question is quite unlike ourselves. Falsehoods are not, in the relevant sense, feasibly knowable.

At this point I want to observe that Kant himself sometimes seems to employ just such a factive conception of possibility, and in a context where he also seems to express a form of semantic anti-realism. For example:

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 (A492-3/B521; cf. A155-6/B194-5, A218-26/B266-73, A495/B523)

If the heart of anti-realism is the claim that meaning and thereby truth bear an essential connection to our cognitive capacities, this certainly looks like a form of anti-realism. Kant seems to suggest here that there being inhabitants

of the moon is a matter of it being possible to experience inhabitants of the moon. This claim involves a factive conception of possibility quite unlike metaphysical possibility, since it might be metaphysically possible to experience inhabitants of the moon even if there are in fact no inhabitants of the moon. Indeed, if Kant's notion of experience here is a notion of a kind of knowledge (see e.g. B147, B165-6, B218, B234, B277), then he is also expressing a direct commitment to a form of knowability in this passage. Kant would seem to be offering us something like the following biconditional: ϕ iff it is possible to know that ϕ . This is an epistemic definition of truth of precisely the sort offered by the anti-realist. The left-to-right direction is a knowability principle. The right-to-left direction corresponds to FACT. Of course Kant's example is not of a transcendental truth and experience, even if it is a form of knowledge, is not a form of a priori knowledge. So this is not yet quite what we are looking for. My point for now is just that Kant sometimes seems to employ a conception of possibility akin to the factive, anti-realist conception, notably in connection with one of his core Critical notions, that of possible experience. Hence the label 'transcendental'.¹⁰

Returning to transcendental possibility itself, then, notice that none of this is to say that we cannot evaluate $\diamond K\phi$ and $\diamond \overset{a}{K}\phi$ in non-actual contexts. We certainly can. Take as a toy example a metaphysically possible world w very like this one except that there are pink elephants. We can sensibly ask whether, in w , it is transcendentially knowable that there are pink elephants. In doing so we are no longer asking about how things are for us now, in the current state of information, since w is not the actual world. But we are asking something very similar and the answer seems clear enough. It is, in w , transcendentially knowable that there are pink elephants. Someone in w would just have go out and find one. This does not contravene the factivity of transcendental possibility any more than being able to reason about metaphysical knowability undermines the factivity of knowledge. Doing so just tells us something about how transcendental possibility interacts with metaphysical possibility. In this case it tells us that, in light of FACT, $\diamond \diamond K\phi$ entails $\diamond \phi$: if it is metaphysically possible that it is transcendentially knowable

¹⁰ For further discussion of this passage and the factive conception of possibility involved in Kant's conception of possible experience more generally, see Milmed (1967), Allais (2015, 137ff.), Stephenson (2015; 2018), and Gomes & Stephenson (2016: 73f.).

that ϕ , then since transcendental knowability is factive, it must be metaphysically possible that ϕ .¹¹ We will see later that being able to reason like this will also be important for understanding both our candidate knowability principle and the result of our new knowability proof. For the moment let us stick with the relation between the transcendental and the metaphysical modalities.

Given their differences when it comes to factivity, it should be unsurprising that transcendental knowability is formally stronger than metaphysical knowability. $\diamond K^a\phi$ entails $\diamond K\phi$ and $\diamond K\phi$ entails $\diamond K^a\phi$, but not conversely. This will be important in what follows and so is worth elaborating on. (I consider the general case but the same holds, mutatis mutandis, for the a priori case).

Transcendental knowability entails metaphysical knowability. I said above that we can think of the factivity of the transcendental knowability operators as resulting from a quasi-stipulative constraint on what is meant by ‘feasibly knowable’ in the anti-realist context. The current claim can well be thought of as a much weaker constraint on what such talk can mean. We have seen that the truth of $\diamond K\phi$ is supposed to be compatible with us in fact never coming to know that ϕ . The claim is just that it would be feasible for us to carry out the procedures required to come to know that ϕ . So either we do at some point carry out these procedures and so come to know that ϕ , or we never carry out these procedures and so never come to know that ϕ . If the former, then the actual world is itself a world in which $K\phi$ is true, and $\diamond K\phi$ follows since the actual world is a metaphysically possible world. If the latter, then although the actual world is not itself a world in which $K\phi$ is true, $\diamond K\phi$ implies that this is only because it just so happens that no one ever carries out the necessary procedures in the actual world. So there is surely a metaphysically possible world just like the actual world except that someone *does* happen to carry out those procedures and so comes to know that ϕ . Either way, $\diamond K\phi$ holds whenever $\diamond K^a\phi$ does.

¹¹ Necessitating on FACT, this is just a special case of normal modal closure—see below.

The converse does not hold, however. That there is some possible world metaphysically accessible from the actual world in which it is known that ϕ does not entail that we could feasibly come to know ϕ . If ϕ is a contingent matter then perhaps ϕ is just false in the actual world, and so not feasibly knowable by the contrapositive of FACT. And even if ϕ is necessary and so true in the actual world, there might still be necessary conditions on our coming to know ϕ that are themselves contingent and fail to actually obtain. Consider, for example, the necessary truth that all red things are colored things. It is a necessary condition on us knowing such a truth that we have the concept <red>. Suppose further that it is a necessary condition on us having the concept <red> that red things exist, for instance because we need to see something red if we are to grasp the concept <red> (see e.g. *Anthropology* 7:168). Now suppose that our world is without red things. (We are reasoning in a non-actual context again.) Then it would not be feasible—it would not be possible in our transcendental, anti-realist sense—for us to know that all red things are colored things. Yet plausibly it would still be metaphysically possible to know that all red things are colored things so long as some worlds metaphysically accessible from our redless world are not themselves redless. Transcendental knowability can fail when metaphysical knowability holds. The latter does not entail the former.¹²

Finally, consider the following closure principle (where the box is the familiar metaphysical necessity operator):

$$(\text{CLOS}) \ \diamond K\phi \wedge \Box(K\phi \rightarrow K\psi) \rightarrow \diamond K\psi$$

CLOS says that if it is feasible for someone to come to know a priori that ϕ and if every metaphysically possible world in which someone knows a priori that ϕ is also a world in which someone knows (in whatever way) that ψ , then it had better be feasible for someone to come to know ψ too (in whatever way). This seems very reasonable. It is important to be clear that, despite a

¹² In these ways, including factivity, the notion of ‘being feasibly knowable for a subject’ is similar to that of ‘a subject’s being in a position to know’—see e.g. Williamson (2000, 94-5). Yet transcendental knowability is weaker. It allows us to be significantly further away from actual knowledge than does the notion of being in a position to know. We might usefully think of transcendental knowability as lying somewhere modally between being in a position to know and metaphysical knowability. See §8 for further discussion.

superficial notational similarity and unlike its metaphysical knowability counterpart, CLOS is *not* an instance of the schema $\Diamond\phi \wedge \Box(\phi \rightarrow \psi) \rightarrow \Diamond\psi$, which holds in any normal modal logic. For as we have seen, \Diamond is quite different from \Box and so, in particular, is not the dual of \Box . Nevertheless, the intuitive plausibility of the normal schema carries over to CLOS. For in what sense could it be feasible for someone to know a priori that ϕ if it were not likewise feasible for someone to know something the knowing of which is a necessary condition of someone knowing a priori that ϕ ? The reasoning here is the same as above. Certain necessary conditions on knowledge will transpose into necessary conditions on transcendental knowability.¹³

Like FACT, CLOS will be central to our knowability proof. Sven Rosenkranz (2004) has argued that a schema closely related to CLOS cannot be jointly valid with FACT. I will respond to this objection in §7. In short, it does not apply to CLOS itself. For now let us turn to our candidate knowability principle, and in particular the question of whether we should attribute it to Kant.

3. The completability of metaphysics

Is Kant an anti-realist about transcendental truths in the sense that he thinks all such truths bear an essential connection to our cognitive capacities, and if so, does he therefore think that all transcendental truths are transcendentially knowable a priori? Where T is our transcendental truth predicate, ‘It is a transcendental truth that’, we can make an initial pass at our candidate knowability principle:

$$(KPT) T\phi \rightarrow \Diamond \overset{a}{K}\phi$$

In an actual context, KPT says that if ϕ is a transcendental truth, then given how things are with us now, in the current state of information, it would be

¹³ We might even think that the transcendental knowability operators are normal modal operators that function like the box. I remain neutral on this here but it does not seem wildly implausible that, on certain idealizations, they would obey the K-axiom: in the general case, that $\Diamond K(\phi \rightarrow \psi) \rightarrow (\Diamond K\phi \rightarrow \Diamond K\psi)$. If so, CLOS is not very far away. And since we already have FACT corresponding to the T-axiom, obeying the K-axiom would mean that, like (idealized) knowledge itself on standard views, transcendental knowability could be modeled in the system \mathcal{KT} . All of this provides further support for CLOS.

feasible for someone to perform procedures such that they come to know a priori that ϕ .¹⁴ But what about non-actual contexts? Plausibly, we need to restrict the contexts in which KPT can be applied. For as with the example from the previous section concerning knowledge that all red things are colored things, there would seem to be contingent conditions on it being feasible for us to come to know transcendental truths.

First, we must exist. If we do not exist, then it would not be feasible, in the transcendental, anti-realist sense, for us to come to know anything. Otherwise put, it is plausible that $\diamond K^a\phi$ is false whenever humans do not exist. Without some restriction on KPT, it would follow by contraposition that $T\phi$ is false whenever humans do not exist. This might be expected on certain strong readings of Kant's idealism. After all, transcendental truths are in *some* sense truths about us. (More on this in a moment.) And of course Kant does at one point say 'if I were to take away the thinking subject, the whole corporeal world would have to disappear' (A383).¹⁵ But restricting the application of KPT to contexts in which humans exist ensures that it is compatible with a much broader range of interpretations of Kant's idealism.

Second, if we are to count as transcendental truths the principles from the *Metaphysical Foundations*, then there must also be matter so that we could feasibly acquire the concept <matter>.¹⁶ This is like the example of <red> from the previous section except that it is much more general. Possession of this extremely basic but nevertheless *empirical* concept is a necessary condition on knowing, for example, that 'Every change in matter has an external cause' (4:543).

¹⁴ Note that KPT entails $T\phi \rightarrow \diamond K\phi$ since knowing a priori is a way of knowing, while $T\phi \rightarrow \diamond K\phi$ does not entail KPT, since knowing a priori is not the only way of knowing. Strictly, then, KPT is the stronger claim, though the difference is insubstantial in this context. If Kant does think that all transcendental truths are transcendently knowable, then the kind of knowledge he thinks it feasible for us to acquire will be a priori knowledge. Nevertheless, it is important that we work with KPT because its extra formal strength allows CLOS and the luminosity principle introduced in the next section to be correspondingly weaker. See also §§7-8.

¹⁵ Interestingly, there is a striking passage in Dummett (2004, 92) that also seems to imply something like this view.

¹⁶ The second condition is distinct from the first, though plausibly they are not independent—that depends on whether humans must be materially embodied to exist.

These are important restrictions on when we can apply our candidate knowability principle, and they will percolate through to the conclusion of any proof in which it is invoked. Thus, for instance, the conclusion of our new knowability proof will only be that all transcendental truths are known whenever humans and matter exist. Even so, this remains a remarkable conclusion.

Building the restrictions into our candidate knowability principle:

(KPT*) $T\phi \rightarrow \diamond K^a\phi$, whenever humans and matter exist

KPT* says that if ϕ is a transcendental truth, then it would be feasible for someone to perform procedures such that they come to know a priori that ϕ in any situation in which humans and matter exist.

It strikes me as eminently plausible to attribute KPT* to Kant. Clearly Kant thinks that transcendental truths bear an essential connection to our cognitive capacities. As he says in the Preface to the *Metaphysical Foundations*, ‘All true metaphysics is drawn from the essence of the faculty of thinking itself’ (4:472). And a year later in the Preface to the B-edition *Critique*, ‘we can cognize of things a priori only what we ourselves have put into them’ (Bxviii, cf. Bxxiii). This much is involved in the Copernican hypothesis and the Critical turn, whatever one makes of the idealism that results. And in each case, Kant goes on to connect this view to knowability via an insistence that (true, scientific) metaphysics is essentially completable:

In everything that is called metaphysics one can hope for the *absolute completeness* of the sciences, of such a kind one may expect in no other type of cognition. Therefore, just as in the metaphysics of nature in general, here also the completeness of the metaphysics of corporeal nature can confidently be expected. The reason is that in metaphysics the object is only considered in accordance with the general laws of thought (4:473)

For that this should be possible, indeed that such a system [‘the philosophy of pure reason’] should not be too great in scope for us to hope to be able entirely to complete it, can be assessed in advance from the fact that our object is not the nature of things, which is

inexhaustible, but the understanding, which judges about the nature of things, and this in turn only in regard to its a priori cognition, the supply of which, since we do not need to search for it externally, *cannot remain hidden from us* (A12-13/B26, my emphasis; cf. Axx, Bxxiv)

In these passages Kant shows a clear commitment to some form of knowability principle for transcendental truths. The only question is, why think it is of the transcendental kind involved in KPT*, over and above a mere metaphysical knowability principle?

First of all note that Kant suggests that transcendental truths *cannot* remain hidden from us (see also Axx). It would suffice for metaphysical knowability that transcendental truths *might* not remain hidden from us—that they are *not necessarily* hidden from us. But what Kant says is much stronger, namely that transcendental truths are *necessarily not* hidden from us. This is not captured by any of the following claims: that transcendental truths are metaphysically knowable; that transcendental truths are necessarily metaphysically knowable; that, necessarily, transcendental truths are metaphysically knowable. For each of these could be true even if there were contingent conditions on being able to come to know transcendental truths that failed in some worlds, including the actual world. To capture Kant's position here without assuming straight off that Kant thinks metaphysics as a system of knowledge is somehow always already complete, we need an account of knowability *within* a world rather than across worlds. The situation-theoretic semantics of transcendental knowability provides just this.

Kant's idea in the above passages seems to be that the particular contingencies of the world in which we find ourselves are of little importance when it comes to being able to know transcendental truths, since in knowing such truths we are really only concerned with ourselves and very general features of the world.¹⁷ Now, plausibly, Kant means certain minimal conditions to be implicit in this, namely that humans and matter exist. Once these conditions are satisfied, however, the idea would be that there is nothing further in

¹⁷ Of course the contingencies of the world in which we find ourselves *should* matter very much for whether we do *in fact* go on to complete metaphysics. This is why the result of our new knowability proof is so surprising.

principle to bar us from being able to know any transcendental truth, and thus, ultimately, from being able to complete metaphysics. This is precisely the picture captured by KPT*.

Another way to think of this is in terms of a material collapse of the formal distinction between metaphysical and transcendental knowability. We saw in §2 that transcendental knowability entails metaphysical knowability but not conversely because the latter can hold when the former fails. There were two kinds of case in which such a mismatch might occur. First, when dealing with contingent propositions that happen to be false in the context of evaluation. Second, when there are contingent conditions on knowledge that fail to hold in the context of evaluation. The first kind of case is irrelevant when it comes to our candidate knowability principle because Kant thinks all transcendental truths are necessary truths (e.g. B3-4). And the second kind of case is effectively ruled out by the restrictions appended to KPT to yield KPT*. Although transcendental and metaphysical knowability are not *in general* equivalent, Kant's views on the necessity of transcendental truths and the restrictions involved in KPT* render the two kinds of knowability equivalent for the cases at hand.

We saw in §2 that Kant sometimes employs a conception of possibility that is much closer to our transcendental conception than to the standard metaphysical conception. There is much in the above passages to indicate that he does so here as well. But ultimately it does not matter which specific conception of possibility he has in mind. The view Kant expresses in these passages is highly suggestive a commitment to KPT* regardless. I propose, then, that there very good grounds for attributing KPT* to Kant.

4. The luminosity of a priori knowledge

We need one more principle before we can construct our new knowability proof. Unlike the old knowability proof, the new proof requires a kind of luminosity principle concerning a priori knowledge. It is natural for anti-

realists to validate such principles,¹⁸ but as we'll see, the details matter, and we need to say a little more. Consider first the following full iteration principle:

$$(KK) \Box (\overset{a}{K}\phi \rightarrow \overset{a}{K}\overset{a}{K}\phi).$$

Necessarily, if it is known a priori that ϕ , then it is known a priori that it is known a priori that ϕ . KK is highly implausible in the Kantian context, for Kant holds that a priori knowledge is knowledge of necessary truths, so KK would entail that it is necessary that it is known a priori that ϕ whenever it is known a priori that ϕ . But this is surely false. Consider again an example from §2. Suppose that a subject S knows a priori that all red things are colored things. This knowledge is contingent on there being red things, for otherwise S would not have the concept <red>, and without this concept, S would not know that all red things are colored things. It is not necessary that there are red things, so it is not necessary that S knows that all red things are colored things. A similar point can be made about other contingent conditions on S having a priori knowledge, for instance that S exists. And more generally, whether or not we happen to acquire concepts, undertake transcendental philosophy, prove theorems in mathematics, and so forth, is surely a contingent matter.

Moreover, if we were to add the B-axiom to our system for metaphysical modality, then an absurdly strong principle like KK—given Kant's views on a priori knowledge and necessity—would yield a suspiciously quick proof of omniscience from knowability. Assuming a merely *metaphysical* knowability principle for transcendental truths, $T\phi \rightarrow \Diamond K\phi$, as well a merely *normal* modal closure principle, $\Diamond\phi \wedge \Box(\phi \rightarrow \psi) \rightarrow \Diamond\psi$, KK would allow us to derive $T\phi \rightarrow \Diamond\Box\overset{a}{K}\phi$, which by the B-axiom entails $T\phi \rightarrow \overset{a}{K}\phi$.

Finally, consider that, as a species of representation, a priori knowledge is a mental state, a 'modification of the subject or soul' (A34/B50, A50/B74, A197/B242). And plausibly, for Kant, our knowledge of what particular mental states we are in is had (at least in part) through inner sense and thus

¹⁸ See Dummett (1981, 632; 1993, 4), Tennant (1997, 43, 203), Brogaard and Salerno (2002, 145), Rosenkranz (2004, 69), and Shapiro (2006, 24-5).

is itself empirical knowledge, even when the states in question are themselves states of a priori knowledge.

All of the above suggests that we at least need to weaken our luminosity principle, like so:

$$(KK^*) \square (\overset{a}{K}\phi \rightarrow K\overset{a}{K}\phi)$$

Necessarily, if it is known a priori that ϕ , then it is known, *in some way*, that it is known a priori that ϕ . But why attribute such a luminosity principle to Kant at all, even if it avoids the above objections? The basis for attributing KK^* to Kant is the specific connection he draws between a priori knowledge and complete or apodictic certainty. He says, for instance:

Reason in abstraction from all experience can know everything only a priori and necessarily, or not at all; hence its judgment is never an opinion, but either abstention from all judgment or apodictic certainty (A775/B803; cf. A822-2/B850-1)

This suggests a very demanding conception of a priori knowledge on which the argument for an epistemic luminosity principle is familiar. Take the general case and suppose that knowledge requires certainty in the sense that it requires consciously ruling out all possibility of error. First, in *consciously* ruling out all possibility of error, this would seem to involve believing that one knows. Second, in consciously *ruling out all possibility of error*, it would seem to involve ruling out the possibility that one does not know, and thus ruling out the possibility that one's belief that one knows is false. Together, this seems to imply that knowing involves knowing that one knows, if knowledge requires certainty. It is worth quoting at length from a famous passage from Jaakko Hintikka (1970, 145-6) on the point:

Suppose we say that evidence for a proposition, P, is *conclusive* iff it is so strong that, once one discovers it, further inquiry cannot give one reason to stop believing P. The concept of knowledge used by many philosophers seems to be a strong one on which one knows P only if one's evidence for P is conclusive in this sense. It is plausible that the KK principle holds for this strong concept of knowledge. For it is plausible that one's evidence for P is conclusive in the above

sense only if it rules out the possibility that one does not know P, and thus only if it allows one to know that one knows P.

To see this, suppose one has evidence, E, for a proposition P, and that E does not rule out the possibility that one does not know P. If E does not rule out this possibility, then, after one has discovered E, further inquiry can, in principle, reveal to one that one does not know P. But if further inquiry were to reveal this, then it would surely give one reason to stop believing P (since one should not believe things that one does not know). So it is plausible that, if E does not rule out the possibility that one does not know P, then it is not conclusive in the sense just defined, and hence plausible that, if knowledge requires evidence that is conclusive in this sense, the KK principle holds.

Placing a strong certainty condition on knowledge provides the original and natural home for epistemic luminosity principles. And in addition to the above reasoning, this allows such principles to avoid the common objection that epistemic standards go up with each iteration of knowledge such that eventually it must be conceded that one does not *know* that one knows that knows that one knows.... If we start straight off with the highest possible epistemic standards, this kind of objection fails to gain a foothold.¹⁹ Though note that in any case, KK* is not itself an iteration principle, so it also avoids another common objection that related, iteratable versions face when attended by a mental state conception of knowledge, namely that they attribute to a subject an increasingly implausible number of mental states of increasingly implausible complexity. To be sure, KK* doubles up on what we know when we know a priori, but it goes no further.

Altogether I think this makes a decent *initial* case that Kant is committed to KK*. There were overriding reasons for Kant not to adopt KK, but these don't apply to KK*. And Kant sees a constitutive connection between a priori knowledge and certainty, specifically that if one lacks certainty then one lacks a priori knowledge. KK* is a *prima facie* plausible way to cash out this part of his view. Otherwise put, it would certainly be a surprising result if we find that Kant ultimately has to reject KK* and the luminosity of a priori knowledge. There is more to be said here, but having made my initial case, I postpone further discussion until §8.

¹⁹ See DeRose (2002).

5. Out with the old, in with the new

We are finally in a position to construct our new knowability proof.²⁰ But first I want to explain a little more fully than I did in §1 why all of the preceding has been necessary; that is, I want to explain why KPT* is restricted enough to avoid the old knowability proof. That proof turns on the unknowability of statements of the form $\phi \wedge \neg K\phi$. Since such statements are provably unknowable, an unrestricted knowability principle that covers all truths whatsoever entails that no such statement is a truth, which is to say that there are no unknown truths. Restricting our knowability principle to transcendental truths quickly yields the corresponding result that no such statement is a transcendental truth, or $\neg T(\phi \wedge \neg K\phi)$, since if any such statement $\phi \wedge \neg K\phi$ were a transcendental truth, it would by the restricted principle be knowable, which no such statement is. However, unlike the general case, it would only follow that there are no unknown transcendental truths on the additional assumption that $T\phi \wedge \neg K\phi$ entails $T(\phi \wedge \neg K\phi)$, so that $\neg T(\phi \wedge \neg K\phi)$ entails $\neg(T\phi \wedge \neg K\phi)$ and thus that there are no unknown transcendental truths. But I can see no reason to think this holds.²¹ And so long as no such entailment holds, we are left only with the result that no statement of the form $\phi \wedge \neg K\phi$ is a transcendental truth.

Far from being problematic, that no statement of the form $\phi \wedge \neg K\phi$ is a transcendental truth is just what is required to avoid the old knowability proof, which only goes through if we can substitute some such statement into our candidate knowability principle. This is why we need a new knowability proof, one that does not require such a substitution. And developing the toolbox for such a proof has been the purpose of the preceding sections.

I restate the relevant principles for ease of reference:

(KPT*) $T\phi \rightarrow \diamond^a K\phi$, whenever humans and matter exist

(KK*) $\Box(K\phi \rightarrow K^a K\phi)$

²⁰ See Williamson (1992) and Brogaard and Salerno (2002) for related arguments.

²¹ It is precisely at this point where matters are different for empirical truths, so that a version of the old knowability proof can be adapted to that case. See Stephenson (2015).

$$\begin{aligned} & \text{(CLOS)} \quad \diamond \overset{a}{K}\phi \wedge \square (\overset{a}{K}\phi \rightarrow K\psi) \rightarrow \diamond K\psi \\ & \text{(FACT)} \quad \diamond K\phi \rightarrow \phi \end{aligned}$$

Our new knowability proof then runs as follows:

- | | |
|--|---|
| (1) $T\phi \wedge \neg K\phi$ | assumption for reductio |
| (2) $\diamond \overset{a}{K}\phi$ | left conjunct of 1, KPT* |
| (3) $\square (\overset{a}{K}\phi \rightarrow K\overset{a}{K}\phi)$ | KK* |
| (4) $\diamond K\overset{a}{K}\phi$ | 2, 3, CLOS |
| (5) $\overset{a}{K}\phi$ | 4, FACT |
| (6) $K\phi$ | 5, knowing a priori is a way of knowing |
| (7) $K\phi \wedge \neg K\phi$ | right conjunct of 1, 6 |
| (8) $\neg(T\phi \wedge \neg K\phi)$ | 1-7, discharging assumption at 1 |

I have not defended or formalized the principle that takes us from (5) to (6) because it is trivial. Since the step from (1) to (2) is only valid when humans and matter exist, we should understand our result at (8) as the claim that, whenever humans and matter exist, there are no unknown transcendental truths. Note that the result would therefore not be accounted for by Kant taking himself to know all transcendental truths. It holds for arbitrary knowers at arbitrary times, so long as humans and matter exist. If the preceding is right, then Kant is committed to the thesis holding even if he had never awoken from his dogmatic slumber, if metaphysics as a discipline had ended with Hume or had never got going in the first place.

Does this amount to a reductio of Kant's position? In the remainder of this paper I will consider ways of avoiding that conclusion. I have made an initial case for each of the principles involved in our new knowability proof, but is there more to say? I take KPT* to be secure, but in §7 I will return to CLOS and FACT and in §8 I will return to KK* and the luminosity of a priori knowledge. First, I want to explain why Kant cannot simply embrace the result that there are no unknown transcendental truths whenever humans and matter exist.

6. A priori cognition as tacit knowledge

It is a point now well recognized that Kant's conception of cognition (*Erkenntnis*) may be fundamentally different from our contemporary conception of knowledge.²² The point goes beyond observing an infelicity of translation. One of Kant's primary concerns in the *Critique of Pure Reason* is an account of a priori cognition in metaphysics. If this is not a theory of a priori knowledge, then what is it?

One interesting answer here is that Kant's account of a priori cognition in metaphysics is an account of what we must all always already count as knowing *tacitly* if we are to have a world in view at all.²³ As he says in the B-edition Introduction, 'We are in possession of certain a priori cognitions, and even the common understanding is never without them' (B3). In the *Metaphysical Foundations of Natural Science*: 'Hence all natural philosophers who have wished to proceed mathematically in their occupation have always, and must have always, made use of metaphysical principles (albeit unconsciously), even if they themselves solemnly guarded against all claims of metaphysics upon their science.' (4:472). And he is recorded as putting the point even more clearly in his mid-1780s lectures on anthropology: 'For all propositions of philosophy are known to everyone... Hence Socrates rightly says: "I am not the teacher of my listeners, but only the midwife of their thoughts"' (25:1221-2).

On this interpretation, it is then one of the central tasks of transcendental philosophy to make such tacit knowledge explicit. As Kant puts it in introducing the Principles, 'Now our task is to *exhibit* in systematic combination the judgments that the understanding *actually* brings about a priori' (A148/B187, my emphasis). This constitutes a move from *Erkenntnis* to *Wissen*, from mere cognition to scientific knowledge proper.

I think there is something importantly right about this picture. It provides an interesting perspective on Kant's Critical project and it coheres with a number

²² See e.g. Gomes & Stephenson (2016), Watkins & Willaschek (2017), Schafer (forthcoming), and Kohl (forthcoming).

²³ See Kitcher (2006).

of Kant's core doctrines, including those in his practical philosophy. But it won't help Kant here.

The problem is that our new knowability proof can easily be adapted to yield the result that there are no transcendental truths that are not *explicitly* known (by arbitrary knowers at arbitrary times whenever humans and matter exist). When Kant talks about the essential completability of metaphysics as a science, he is clearly talking about the possibility of attaining explicit knowledge. Explicit knowledge, after all, is what constitutes a science. Our closure and factivity principles for transcendental possibility were motivated by considerations independent of any tacit/explicit distinction. And a central motivation for our luminosity principle for a priori knowledge depended on considerations of explicit, reflective certainty. If Kant holds such a luminosity principle at all, then he holds it for explicit a priori knowledge.

All of this means that we can—indeed, *should*—think of our knowledge operators as operators for explicit knowledge. Where ϕ is a transcendental truth and humans and matter exist, it is transcendently possible for us to have explicit a priori knowledge that ϕ (KPT*). Necessarily, whenever we have explicit a priori knowledge that ϕ , then we know explicitly that we have explicit a priori knowledge that ϕ (KK*). And so on. The conclusion of our proof would then be the claim that, whenever humans and matter exist, there are no transcendental truths that are not explicitly known by some arbitrary knower at some arbitrary time. Regardless of Kant's views on a priori cognition as tacit knowledge, indeed regardless of his idealism, this just seems an implausible claim. Our new knowability proof threatens a reductio of Kant's position. If he is to avoid it, he has to find fault with one of the principles that is involved in the proof itself.

7. Factivity and closure again

Our new knowability proof requires FACT and CLOS, which I restate for ease of reference:

$$\text{(FACT)} \quad \diamond K\phi \rightarrow \phi$$

$$\text{(CLOS)} \quad \diamond \overset{a}{K}\phi \wedge \square (\overset{a}{K}\phi \rightarrow K\psi) \rightarrow \diamond K\psi$$

FACT says that transcendental knowability is factive. CLOS says that, if it is transcendently possible for someone to come to know a priori that ϕ , and if every metaphysically possible world in which someone knows a priori that ϕ is also a world in which someone knows (in whatever way) that ψ , then it had better be transcendently possible for someone to come to know ψ too (in whatever way).

Rosenkranz (2004) has argued that a principle closely related to CLOS cannot be jointly valid with FACT. If Rosenkranz's objection is sound and can be applied to CLOS itself, then our new knowability proof would fail and Kant would be out of the woods. Rosenkranz's target schema is the following generalization of CLOS:

$$(\text{CLOS}^*) \ \diamond K\phi \wedge \Box(K\phi \rightarrow K\psi) \rightarrow \diamond K\psi$$

CLOS* says that, if it is transcendently possible for someone to come to know (*in whatever way*) that ϕ , and if every metaphysically possible world in which someone knows (*in whatever way*) that ϕ is also a world in which someone knows (in whatever way) that ψ , then it had better be transcendently possible for someone to come to know ψ too (in whatever way).

Rosenkranz argues that FACT and CLOS* are incompatible as follows. Suppose that the pharaoh is in the tomb and that, *as a matter of contingent fact*, coming to know that the pharaoh is in the tomb would involve coming to know that the door of the tomb has been opened. Where ϕ is 'The pharaoh is in the tomb' and ψ 'The door of the tomb has been opened': $\phi \wedge (K\phi \rightarrow K\psi)$. Now suppose further that it is transcendently possible for someone to come to know this, i.e. $\diamond K(\phi \wedge (K\phi \rightarrow K\psi))$. Finally, since knowledge distributes over conjunction and is factive, we can assume the following: $\Box(K(\phi \wedge (K\phi \rightarrow K\psi)) \rightarrow K\psi)$. That is, it is metaphysically necessary that knowing $\phi \wedge (K\phi \rightarrow K\psi)$ implies knowing ψ . CLOS* now tells us that ψ is transcendently knowable, or $\diamond K\psi$, in which case FACT tells us that ψ is actually true—the door of the tomb has in fact been opened. But why should we think that the door of the tomb has in fact been opened? As Rosenkranz (2004, 71) puts it, 'the door may well not have been opened. To be sure, our

assumptions entail that, once we actually come to know $[\phi]$, $[\psi]$ will be true. But that $['K\phi']$ is true was not one of our assumptions.'

We can grant for the sake of argument that this objection is right as far as it goes.²⁴ But does it apply to CLOS?

First of all note that CLOS does *not* entail CLOS*. For $\diamond \overset{a}{K}\phi$ entails $\diamond K\phi$ but not conversely—being transcendentally knowable a priori is but one way among others of being transcendentally knowable. CLOS in effect strengthens the antecedent of CLOS* so that CLOS is the weaker principle. Thus Rosenkranz's objection to CLOS* has no immediate consequences for CLOS. Unless Rosenkranz's objection can be applied directly to CLOS, the principle can be allowed to stand alongside FACT.

Here is what Rosenkranz's premises would need to look like were we to substitute CLOS* with CLOS in his reasoning:

- (1) $\phi \wedge (\overset{a}{K}\phi \rightarrow K\psi)$
- (2) $\diamond \overset{a}{K}(\phi \wedge (\overset{a}{K}\phi \rightarrow K\psi))$
- (3) $\square(\overset{a}{K}(\phi \wedge (\overset{a}{K}\phi \rightarrow K\psi)) \rightarrow K\psi)$

The parallel objection would then be that FACT and CLOS allow us derive ψ from these premises, but we shouldn't be able to derive ψ from these premises, so FACT and CLOS cannot be jointly valid. But in fact, in this case and unlike Rosenkranz's original case, we should not be at all surprised that we can derive ψ from (1)-(3), nor is it in any way problematic that we can. It turns out that the additional apriority indexes are crucial to avoiding Rosenkranz's objection, for Kant thinks that a priori knowledge is knowledge of necessary truths. I will explain

The substantial premise here is (2). (1) follows from (2) by the factivity of transcendental knowability and (3) is simply an application of the factivity and distributivity over conjunction of knowledge. So let us look a little more closely at (2).

²⁴ Though see Brogaard and Salerno (2006) for a response.

Transcendental a priori knowability also distributes over conjunction. How could it be feasible to know a conjunction a priori unless it were feasible to know each of its conjuncts a priori? Thus (2) entails:

$$(2^*) \diamond K^a \phi \wedge \diamond K^a (K^a \phi \rightarrow K \psi)$$

Kant holds that a priori knowledge is knowledge of necessary truths, so, in the present context, we can take it that (2^{*}) in turn entails:

$$(2^{**}) \diamond K^a \phi \wedge \square (K^a \phi \rightarrow K \psi)$$

And here we come to the crucial difference between the two cases. For it was important for Rosenkranz that it was merely a contingent matter of fact that coming to know ϕ happened to involve coming to know ψ , and this is no longer the case. The right conjunct of (2^{**}) says that knowing ψ is *necessary* for knowing ϕ . And since the left conjunct says that it is transcendently possible for someone to come to know a priori that ϕ , we are in effect back to the intuitive plausibility behind CLOS. If knowledge of ψ is metaphysically necessary for a priori knowledge of ϕ , then it better be transcendently possible for someone to come to know that ψ if it is to be transcendently possible for someone to come to know a priori that ϕ . And for it to be transcendently possible for someone to come to know ψ , ψ must be true by FACT. If (2) entails (2^{**}), given Kant's conception of a priori knowledge and necessity, and if (2^{**}) entails ψ by exactly the same reasoning that supported CLOS and FACT in the first place, then it is no surprise that applying FACT and CLOS to premises that include (2) should allow us to derive ψ .

Generally speaking, the lesson here is simply that we should not worry about being able to derive statements the truth of which is a necessary condition of our premises. This, after all, is just the structure of transcendental argumentation.

Consider again an example from §2. Let ϕ be 'All red things are colored things' and ψ 'There exist red things'. The latter is presumably contingent. But we should not worry about being able to derive it from premises that include the

claim that it is transcendently possible for someone to come to know a priori that all red things are colored things, since the contingent claim is a *necessary* condition of this. Otherwise put, if no red things exist so that we cannot acquire the concept <red>, then we would be happy to concede that it is not transcendently knowable that all red things are colored things. This was the purpose of our restrictions on KPT* so that Kant's transcendental knowability principle for transcendental truths only holds under the conditions that humans and matter exist. It will come up again in the next section.

If we substitute CLOS* for the weaker principle CLOS, and if we bear in mind Kant's views about a priori knowledge and necessity, then Rosenkranz's reasoning in effect involves premises that *should* entail the statement that is then formally derived. In the Kantian context, CLOS and FACT can stand together.

8. Luminosity reconsidered

This leaves us with just one place to query our new knowability proof. In §4 I appealed to Kant's account of the relation between certainty and a priori knowledge to motivate attributing to him the following luminosity principle:

$$(KK^*) \quad \Box(\overset{a}{K}\phi \rightarrow K\overset{a}{K}\phi)$$

I explained there how KK* is far more defensible than the full iteration principle for a priori knowledge, KK: $\Box(\overset{a}{K}\phi \rightarrow \overset{a}{K}\overset{a}{K}\phi)$. But one may worry that KK* is still too strong. It is common to weaken such principles by appealing to some notion of being in a position to know. If one knows, then one needn't *actually* know that one knows, but one must at least be *in a position to* know that one knows. Before considering the motivation for such principles, let's see how the modification would affect our proof.

Instead of introducing yet another operator for 'being in a position to know', I will use our familiar transcendental knowability operator. I observed in §2 that the notion of a subject's being in a position to know is in many ways similar to the anti-realist's notion of being feasibly knowable for a subject. Both notions are syncategorematic and factive, for instance. And while

transcendental knowability allows us to be significantly further away from actual knowledge than does the notion of being in a position to know, this difference does not significantly affect the current issue. Consider, then, the following weakened version of KK^* :

$$(KK^*-) \quad \Box(\overset{a}{K}\phi \rightarrow \Diamond KK^a\phi)$$

KK^*- says that, necessarily, if it is known a priori that ϕ , then it is transcendently knowable, in some way, that it is known a priori that ϕ . As it stands such a modification would invalidate our new knowability proof. The reason is that our closure principle CLOS can no longer be applied as required, for KK^*- is different in form to the right conjunct of CLOS's antecedent: $\Box(\overset{a}{K}\phi \rightarrow K\psi)$. However, this can be remedied by strengthening CLOS accordingly, like so:

$$(CLOS+) \quad \Diamond \overset{a}{K}\phi \wedge \Box(\overset{a}{K}\phi \rightarrow \Diamond K\psi) \rightarrow \Diamond K\psi$$

Replacing KK^* with KK^*- and CLOS with CLOS+, the proof from §5 runs exactly as before. The question, then, is whether CLOS+ is still plausible. I think it is. CLOS+ says that, if it is feasible for someone to come to know a priori that ϕ and if every metaphysically possible world in which someone in fact knows a priori that ϕ is also a world in which it is feasible for someone to come to know that ψ , then it must in fact be feasible for someone to come to know that ψ . The intuitive plausibility of CLOS carries over to CLOS+. For in what sense could it be feasible for someone to come to know a priori that ϕ if the necessary conditions on knowing a priori that ϕ are not fulfilled? This is a refrain we have seen time and again.

Where does this leave us? I argued in §4 that Kant's reflective certainty condition on a priori knowledge suggests that he thinks we must *actually* know that we know a priori that ϕ , whenever we know a priori that ϕ . Whence KK^* . And since KK^* entails KK^*- , motivation for KK^* is motivation for KK^*- . But the converse does not hold: KK^*- does not entail KK^* . So the question arises: Are there motivations for attributing to Kant KK^*- only? That is, are there motivations for KK^*- that don't stem from motivations for KK^* ?

There are, and they are worth running through because we might find them more compelling than the previous certainty considerations.

Williamson (2000, 93-4) begins his celebrated anti-luminosity argument with the following reflections:

There is a constant temptation in philosophy to postulate a realm of phenomena in which nothing is hidden from us. Descartes thought that one's own mind is such a realm. Wittgenstein enlarged the realm to everything that is of interest to philosophy. That they explained this special feature in very different ways hardly needs to be said; what is remarkable is their agreement on our possession of a cognitive home in which everything lies open to our view. Much of our thinking—for example, in the physical sciences—must operate outside this home, in alien circumstances. The claim is that not all our thinking could be like that.

To deny that something is hidden is not to assert that we are infallible about it. Mistakes are always possible. There is no limit to the conclusions into which we can be lured by fallacious reasoning and wishful thinking, charismatic gurus and cheap paperbacks. The point is that, in our cognitive home, such mistakes are always rectifiable. Similarly, we are not omniscient about our cognitive home. We may not know the answer to a question simply because the question has never occurred to us. Even if something is open to view, we may not have glanced in that direction. Again, the point is that such ignorance is always removable.

We saw in §3 that Kant thinks metaphysics 'cannot remain hidden from us' (A13/B26; cf. Axx, Bxxiv). His account exemplifies the one sketched by Williamson remarkably closely. Kant does not think we are infallible in metaphysics. Indeed, he thinks that mistakes in the form of transcendental illusions are not only possible but natural (A293/B349, Avii). Nor does he think we are omniscient in metaphysics. We may not know the answer to a question simply because it has never occurred to us, such as how synthetic a priori cognition is possible (B19, A764/B792, A10n.; *Prolegomena* 4:270). But he clearly thinks that metaphysics is our cognitive home, where mistakes are always rectifiable and ignorance is always removable. And the reasoning behind this view is familiar, even if Kant's conception of metaphysics is revolutionary: metaphysics is our cognitive home because metaphysical

knowledge is a kind of *self*-knowledge (Axi, Axx, Bxviii, Bxxiiiif., A13/B26; *Metaphysical Foundations* 4:472-3).

We saw in §3 how these views support attributing to Kant a knowability principle for transcendental truths. And we can now see that the *selfsame* views support attributing to Kant a luminosity principle like KK^{*}-. If metaphysics is our cognitive home such that we can always remove ignorance and error, then it should always be at least feasible for us to know when we have a priori knowledge of transcendental truths.²⁵ Yet these considerations do not motivate KK^{*}-. It should always be feasible for us to know when we have a priori knowledge of transcendental truths, but ignorance and error are possible even in our cognitive home, so we might in fact fail to know when we have a priori knowledge of transcendental truths.

There are different ways to motivate different luminosity principles for a priori knowledge in Kant, and these differences require different commitments regarding the logic of transcendental knowability if our new knowability proof is to work. But whichever route we take, a case can be made for our proof and Kant is in trouble. It may well be that Kant can be defended against the objections I have raised in this paper. But I hope at least to have shown that more work is required on the neglected issues of transcendental knowability and a priori luminosity in Kant.²⁶

²⁵ Strictly speaking, if this is our motivation for luminosity, then we should conditionalize KK^{*}- on $T\phi$, and thus the right conjunct of the antecedent of CLOS+ as well. Certainty is a requirement on all (explicit) a priori knowledge, including mathematical and analytic knowledge, but perhaps it is only metaphysics that is our cognitive home. This would not affect our proof, since we have $T\phi$ in our assumption at line (1), so I ignore the complication here.

²⁶ For discussion and comments, invariably insightful, generous and helpful, my thanks to Catharine Diehl, Max Edwards, Anil Gomes, Nora Kreft, Beau Madison Mount, audiences in Aarhus, Berlin, Halle, Potsdam, Sheffield, and two anonymous referees for this journal.

References

- Allais, L. (2015) *Manifest Reality: Kant's Idealism and his Realism*, Oxford University Press.
- Baldwin, T. (2001) 'Kantian Modality', *Proceedings of the Aristotelian Society*, suppl. 76: 1-24.
- Brogaard, B. and Salerno, J. (2002) 'Clues to the Paradoxes of Knowability: Reply to Dummett and Tennant', *Analysis* 62 (2): 143-50.
- (2006) 'Knowability and a Modal Closure Principle', *American Philosophical Quarterly* 43 (3): 261-70.
- DeRose, K (2002) 'Assertion, Knowledge and Context', *Philosophical Review* 111 (2): 167-203.
- Dummett, M. (1959) 'Truth', *Proceedings of the Aristotelian Society*, 59: 141-162.
- (1981) *The Interpretation of Frege's Philosophy*, Duckworth.
- (1993) *The Seas of Language*, Oxford University Press.
- (2004) *Truth and the Past*, Columbia University Press.
- Gomes, A. & Stephenson, A. (2016) 'On the Relation of Intuition to Cognition', in D. Schulting (ed.), *Kantian Nonconceptualism*, Routledge, 53-79.
- Hintikka, J. (1970) "Knowing that One Knows" reviewed', *Synthese* 21: 141-62.
- Kant, I. (1900-) *Gesammelte Schriften*, De Gruyter and predecessors.
- (1997) *Critique of Pure Reason*, P. Guyer & A. W. Wood (eds. & trans.), Cambridge University Press.
- (2002) *Theoretical Philosophy After 1781*, H. Allison & P. Heath (eds.), G. Hatfield, M. Friedman, H. Allison, & P. Heath (trans.), Cambridge University Press.
- (2007) *Anthropology, History, and Education*, R.B. Louden and G. Zöllner (eds.), M.J. Gregor, P. Guyer, R.B. Louden, H. Wilson, A.W. Wood, G. Zöllner, & A. Zweig (trans.), Cambridge University Press.
- (2012) *Lectures on Anthropology*, A.W. Wood & R.B. Louden (eds.), R.R. Clewis, R.B. Louden, G.F. Munzel, & A.W. Wood (trans.), Cambridge University Press.
- Kitcher, P. (2006) 'A Priori', in P. Guyer (ed.), *Kant and Modern Philosophy*, Cambridge University Press.

- Kohl, M. (forthcoming) 'Kant on Cognition and Knowledge', in A. Gomes & A. Stephenson (eds.), *Oxford Handbook of Kant*, Oxford University Press.
- Melia, J. (1991) 'Anti-Realism Untouched', *Mind* 100 (399): 341-42.
- Milmed, B. (1967) "Possible Experience" and recent interpretations of Kant', *The Monist* 51 (3): 442-62.
- Moore, A. W. (2012) *The Evolution of Modern Metaphysics: Making Sense of Things*, Cambridge University Press.
- Posy, C. (1983). 'Dancing to the Antinomy: A Proposal for Transcendental Idealism', *American Philosophical Quarterly* 20: 81-94.
- Putnam, H. (1981) *Reason, Truth and History*, Cambridge: Cambridge University Press.
- (1987) *The Many Faces of Realism*, LaSalle: Open Court.
- Rosenkranz, S. (2004) 'Fitch Back in Action Again?', *Analysis* 64 (1): 67-71.
- Salerno, J. (2009) *New Essays on the Knowability Paradox* (Ed.), Oxford: Oxford University Press.
- Schafer (forthcoming) 'Kant's Conception of Cognition and our Knowledge of Things in Themselves', in K. Schafer and N. Stang (eds.), *The Sensible and Intelligible Worlds: New Essays on Kant's Metaphysics and Epistemology*, Oxford University Press.
- Shapiro, S. (2006). 'Externalism, Anti-Realism and the KK thesis', in *A Logical Approach to Philosophy* (eds. D. Devidi and T. Kenyon), Dordrecht: Springer, 22-35.
- Stephenson, A. (2015) 'Kant, the Paradox of Knowability, and the Meaning of "Experience"', *Philosophers' Imprint*, 15 (27), 1-19, 2015
- (2018) 'How to Solve the Paradox of Knowability with Transcendental Epistemology', *Synthese*, advanced online publication, doi: 10.1007/s11229-018-1822-8.
- Stevenson (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 University Press.
- (2000). 'Anti-Realist Aporias'. *Mind* 109 (436): 825-54.
- (2002) 'Victor Vanquished', *Analysis* 62: 135-42.

- Walker, R. C. S. (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.
- Watkins, E. & Willaschek, M. (2017) 'Kant on Cognition and Knowledge', *Synthese*.
- Williamson T. (1992) 'On intuitionistic modal epistemic logic', *Journal of Philosophical Logic* 21: 63-89.
- (2000) *Knowledge and its Limits*, Oxford University Press.
- Wright, C. (2001) 'On being in a quandary', *Mind* 110: 45-98.