Synthese (2021) 198 (Suppl 13):S3253–S3278
https://doi.org/10.1007/s11229-018-1822-8

S.I.: THE CURRENT RELEVANCE OF KANT'S METHOD IN PHILOSOPHY

How to solve the knowability paradox with transcendental epistemology

Andrew Stephenson[1]

Received: 28 December 2017 / Accepted: 17 May 2018 / Published online: 9 June 2018

Abstract A novel solution to the knowability paradox is proposed based on Kant's transcendental epistemology. The 'paradox' refers to a simple argument from the moderate claim that all truths are knowable to the extreme claim that all truths are known. It is significant because anti-realists have wanted to maintain knowability but reject omniscience. The core of the proposed solution is to concede realism about epistemic statements while maintaining anti-realism about non-epistemic statements. Transcendental epistemology supports such a view by providing for a sharp distinction between how we come to understand and apply epistemic versus non-epistemic concepts, the former through our capacity for a special kind of reflective self-knowledge Kant calls 'transcendental apperception'. The proposal is a version of restriction strategy: it solves the paradox by restricting the anti-realist's knowability principle. Restriction strategies have been a common response to the paradox but previous versions face serious difficulties: either they result in a knowability principle too weak to do the work anti-realists want it to, or they succumb to modified forms of the paradox, or they are ad hoc. It is argued that restricting knowability to non-epistemic statements by conceding realism about epistemic statements avoids all versions of the paradox, leaves enough for the anti-realist attack on classical logic, and, with the help of transcendental epistemology, is principled in a way that remains compatible with a thoroughly anti-realist outlook.

Keywords Kant · Knowability paradox · Anti-realism · Transcendental epistemology · Transcendental apperception · Self-knowledge · Reflection · Inner sense · Other minds · Dummett

✉ Andrew Stephenson
andrew.stephenson@soton.ac.uk

[1] University of Southampton, Southampton, UK

1 Introduction

The so-called knowability 'paradox' refers to a simple argument from the moderate claim that all truths are knowable to the extreme claim that all truths are known.[1] Whether or not this result marks a genuine paradox, it is certainly surprising. One reason it is significant is that anti-realists have wanted to maintain knowability—the claim that all truths are knowable—but reject omniscience—the claim that all truths are known. If knowability entails omniscience, then such a position is inconsistent.

For the purposes of this paper I follow the standard line in treating the paradox simply as an argument that poses a serious problem for anti-realism, while continuing to refer to it as a paradox. Accordingly, when I talk of solutions to the paradox I mean ways that anti-realism can respond to this problem. Many such solutions have been proposed but perhaps the most prominent has been to restrict the anti-realist's knowability principle in such a way as to avoid the collapse into omniscience. This kind of 'restriction strategy' will be my focus here.

We already have enough to draw an interesting parallel to Kant. It is often thought that anti-realism is a form of transcendental idealism or that Kant is an anti-realist.[2] Yet Kant also restricts knowability in various ways—most famously, of course, he denies that we can have knowledge of things in themselves. It is therefore natural to ask whether there are resources in Kant that are relevant to the issue at hand. It is the aim of this paper to show that there are, and a novel solution to the knowability paradox is proposed based on Kant's transcendental epistemology. What's important here is not Kant's idealism or his humility regarding things in themselves, however, but rather his account of our capacity for a special kind of reflective self-knowledge he calls 'transcendental apperception', of how it differs from receptive self-knowledge through inner sense, and its role in enabling thought about other minds.

In Sect. 2 I introduce anti-realism and present the basic form of the knowability paradox. In Sect. 3 I introduce a novel restriction of knowability to what I call 'non-epistemic' statements and argue that it is preferable to previous restriction strategies in two key respects: it yields a principle strong enough to form the basis of the anti-realist attack on classical logic but weak enough to avoid all versions of the paradox. One of the main challenges for any restriction strategy is to show that the proposed restriction is principled and not just ad hoc. This is the task of Sect. 4, the heart of the paper, where I turn to Kant's transcendental epistemology. Transcendental apperception is our capacity to gain knowledge of the basic rational nature of our own cognitive capacities through exercising those very capacities. I argue that the resultant picture of how we acquire epistemic concepts on the basis of our own epistemic activity, yet apply them to others on an entirely different basis, provides for a principled way in which to concede a strictly limited realism about epistemic statements while maintaining anti-realism

[1] Due to Alonzo Church and Frederic Fitch. See Salerno (2009) and Brogaard and Salerno (2013) for comprehensive overviews and references.

[2] See, e.g., Strawson (1966: p. 16), Putnam (1981: p. 60ff.), Walker (1995), Moore (2012: p. 362ff.), Allais (2015: p. 209ff.), and Stephenson (2015a). For some key passages, see A62/87, A155-6/B194-5, A218-26/B266-73, B279, A492-6/B521-4, A647/B675; *Prolegomena* (4: 290–291, 336–337). References to Kant are to volume and page of the academy edition and are accompanied by a short English title, except those to the *Critique of Pure Reason*, which take the standard A/B format.

about non-epistemic statements. This in turn provides a principled motivation for my proposed restriction of knowability to non-epistemic statements.

2 Anti-realism and the knowability paradox

As the labels will be used here, 'realism' and 'anti-realism' denote views about meaning and truth. The views agree that the meaning of a declarative statement is given by its truth-conditions—how things must stand if the statement is to be true. They disagree about how to understand the notion of truth involved in such a theory of meaning. Anti-realism gives an epistemic characterization of truth such that a statement is true if and only if someone could, at least in principle, come to know it. Realism places no such constraints on truth, holding that a statement can be true independently of whether or not someone could, even in principle, come to know it.

Anti-realism can be captured in the following principle[3]:

$$(\text{AR})\ \phi \leftrightarrow \Delta K\phi$$

K is our epistemic operator. It says 'someone knows, at some time, that'. I will say more in Sects. 3.3 and 4 about what counts as 'someone', including imposing some specifically Kantian constraints. The notion can be left vague for now, except to say that we are not here concerned with divine knowers—it is not in terms of the cognitive capacities of such beings that anti-realism characterizes truth. Δ is our possibility operator. It says 'it could, at least in principle, be the case that'. I will say more about the kind of possibility involved in anti-realism in Sect. 3.2, where we will see that it is quite different from any of the more familiar notions of, say, logical, conceptual, or metaphysical possibility. For the moment all that matters is that anti-realist possibility must be at least as strong as these notions, such that $\Delta K\phi \vdash \Diamond K\phi$ (where $\Diamond$ is your choice of one of the more familiar operators, 'it is logically/conceptually/metaphysically possible that'). This allows us to derive the following knowability principle from the left–right direction of AR:

$$(\text{KP})\ \phi \rightarrow \Diamond K\phi$$

Glossing over the above qualifications: if ϕ, then it's possible to know that ϕ.

I turn to the motivations behind anti-realism in Sect. 4. First let us focus on the knowability paradox. KP is enough to get the basic form of the paradox going. It requires remarkably modest auxiliary principles: that knowledge of a conjunction entails knowledge of each of the conjuncts, that knowledge entails truth, that theorems are necessary, and that necessary falsehoods are impossible. We begin by using these auxiliary principles to show that no statement of the form $\phi \wedge \neg K\phi$ is knowable:

[3] For simplicity I follow common practice and ignore the truth predicate in my formalization of the core principle of anti-realism, assuming that 'ϕ' is true if and only if ϕ. This is not uncontroversial among anti-realists but it won't be important here. For relevant discussion see Murzi (2012: p. 19f.) and Rumfitt (2015: p. 125ff.). I discuss another refinement of the principle in Sect. 4.1.

(1) $K(p \wedge \neg Kp)$	assumption for reductio
(2) $Kp \wedge K\neg Kp$	1, K-DIST: $K(\phi\wedge\psi) \rightarrow K\phi\wedge K\psi$
(3) $Kp \wedge \neg Kp$	2, K-FACT: $K\phi \rightarrow \phi$, on right conjunct
(4) $\neg K(p \wedge \neg Kp)$	1, 3, discharging assumption at 1
(5) $\Box\neg K(p \wedge \neg Kp)$	4, necessitation
(6) $\neg\Diamond K(p \wedge \neg Kp)$	5, modal operator exchange

Now we prove the main result via an application of KP:

(7) $p \wedge \neg Kp$	assumption for reductio
(8) $\Diamond K(p \wedge \neg Kp)$	7, KP
(9) $\neg(p \wedge \neg Kp)$	6, 7, 8, discharging assumption at 7
(10) $p \rightarrow Kp$	9, classical logic
(11) $\forall p\ (p \rightarrow Kp)$	10, universal generalisation

Despite its apparent simplicity, each stage of this little proof has generated considerable discussion. For the purposes of this paper, I assume that the auxiliary principles are all in order and that the proof is valid. I also assume that the omniscience result at line (11) is unacceptable. That leaves us with exactly one place to question whether the proof amounts to a reductio of anti-realism, namely the application of KP to a statement of the form $\phi \wedge \neg K\phi$. This in any case is clearly the heart of the proof. If anti-realism is to avoid collapsing into omniscience, it must restrict its epistemic characterization of truth, and in particular the resultant knowability principle, so that it can no longer be applied to such statements.

This approach to the paradox has been proposed by two of the foremost defenders of anti-realism. Michael Dummett (2001) proposes to restrict anti-realism's epistemic characterization of truth to what he calls 'basic' statements—roughly, statements that are grammatically simple. Neil Tennant (1997) proposes a restriction to what he calls 'Cartesian' statements—statements the knowing of which is not provably inconsistent. Since conjunctions generally are not basic and since statements of the form $\phi \wedge \neg K\phi$ in particular are not Cartesian (i.e. knowing them is provably inconsistent by the first stage of the above proof), both Dummett's and Tennant's restrictions block the paradox.

It is a serious question for any restriction strategy whether it is principled and not just ad hoc. Dummett's and Tennant's proposals have both faced trenchant criticism along these lines. The question of principle can be postponed until Sect. 4, however, because these restriction strategies face more straightforward objections. I will briefly outline these objections (Sects. 3.1, 3.2) before showing how they can be met by an intermediate restriction of knowability to non-epistemic statements (Sect. 3.3). Dummett's and Tennant's proposals lay at opposite ends of a spectrum. Each is too extreme to provide a satisfactory anti-realist solution to the knowability paradox. What the anti-realist needs is something in between.

3 Three restriction strategies

3.1 Basic statements

Start with Dummett's strong restriction of the anti-realist's epistemic characterization of truth to basic statements. It yields the following knowability principle:

$$(\mathrm{KP_B})\ \phi \rightarrow \Delta K\phi, \quad \text{where } \phi \text{ is basic}$$

The problem I want to focus on is that $\mathrm{KP_B}$ is too weak to be able to do the work anti-realists want it to, namely force a rejection of classical logic in favour of intuitionistic logic via what Crispin Wright has called the Basic Revisionary Argument.[4]

Consider the classical law of excluded middle:

$$(LEM)\ \phi \vee \neg\phi$$

Combining LEM with a knowability principle quickly yields a decidability theorem of the following form:

$$(DEC)\ \Diamond K\phi \vee \Diamond K\neg\phi$$

The range of DEC (i.e. the permissible instances of ϕ) will be the intersection of the ranges of LEM and the knowability principle from which DEC was derived. Since classical LEM is unrestricted, the range of DEC will match that of our chosen knowability principle.

Now, suppose our chosen knowability principle is unrestricted so that DEC is unrestricted too. Arguably, this gives the anti-realist reason to reject unrestricted LEM and thus adopt intuitionistic rather than classical logic. The reasoning is roughly as follows. Suppose that our anti-realist takes themself to know their unrestricted knowability principle. If they also took themself to know unrestricted LEM, then they would also take themself to know unrestricted DEC—their claims to knowledge are closed under such a straightforward entailment. But they do not take themself to know something as strong as unrestricted DEC, so they should not take themself to know unrestricted LEM. This is reason to reject unrestricted LEM as a law of logic, as we should only accept as laws of logic those principles that we take ourselves to know. Thus it is reason to reject classical logic in favour of intuitionistic logic.

So why doesn't the anti-realist take themself to know unrestricted DEC? Consider L.E.J. Brouwer's response to David Hilbert's (1902: p. 445) bold (and ill-fated) pronouncement that 'in mathematics there is no *ignorabimus*':

> There is not a shred of a proof for the conviction which has sometimes been put forward that there exist no unsolvable mathematical problems. (Brouwer 1975[1908]: p. 109)

[4] See Wright (1992: p. 37ff.) and Wright (2001: p. 65ff.).

Dummett introduced anti-realism as an extension of Brouwer's intuitionist program beyond the mathematical domain. In doing so, he identified three more general sources of doubt regarding unrestricted DEC, that is, sources of potential undecidability: quantification over infinite totalities (he mentions Goldbach's conjecture and the continuum hypothesis, two of Brouwer's own examples); tense operators (as in 'A city will never be built on this spot'); and counterfactual conditionals (as in 'If Jones had encountered danger, he would have acted bravely').[5]

The problem for Dummett's proposed restriction of knowability to basic statements, then, is that none of his counterexamples to DEC are examples of basic statements, nor do any of the general sources he identifies look especially apt to produce such statements. It is therefore unclear whether Dummett (or anyone else) has provided any reason to doubt a form of DEC *restricted to basic statements*.[6] But if there is no reason to doubt a form of DEC restricted to basic statements, then the Basic Revisionary Argument sketched above will not go through for the anti-realist who restricts knowability to basic statements with $\mathrm{KP_B}$.[7]

This objection to Dummett's proposed restriction strategy for solving the knowability paradox is a serious one. Both the rejection of classical logic in favour of intuitionistic logic and doubt about (suitably unrestricted forms of) DEC are absolutely central to Dummett's anti-realism, as they are to many versions of anti-realism. There may be other routes to intuitionistic logic.[8] Or perhaps anti-realism can be decoupled from the project of logical revision altogether. But for the anti-realist who wants a route to intuitionistic logic via the Basic Revisionary Argument, we have motivation enough to search for an alternative restriction strategy, one that fairs better in this respect.

3.2 Cartesian statements

Let us turn to the other end of the spectrum and Tennant's proposed weak restriction of anti-realism's epistemic characterization of truth to Cartesian statements. Cartesian statements are those statements the knowing of which is not provably inconsistent, i.e. those ϕ such that $K\phi \nvdash \bot$. This restriction yields a knowability principle that cannot be applied to statements of the form $\phi \wedge \neg K\phi$, which are not Cartesian, so it avoids the paradox from Sect. 2. Moreover, such a knowability principle is clearly still strong enough to form the basis of the kind of Basic Revisionary Argument sketched in the

[5] See Dummett (1978: pp. 1–28) and Dummett (1991: p. 315).

[6] In the case of basic arithmetical statements, for instance, we even have a *proof* of decidability, and Dummett himself notes the relevant qualification in this context: 'we cannot, *save for the most elementary statements*, guarantee that we can find either a proof or a disproof of a given statement' (2000: p. 5, my italics).

[7] For versions of this worry, see Tennant (2002: p. 141), Williamson (2009: p. 187), and especially Murzi (2012).

[8] See Rumfitt (2015) for a recent appraisal. It should be noted that the paradox itself cannot provide a route to intuitionistic logic *for the defender of* KP_B. First, the defender of $\mathrm{KP_B}$ takes themself to have solved the paradox already, and so cannot appeal to it to motivate intuitionistic logic. Second, if intuitionistic logic solves the paradox, there was no need to adopt $\mathrm{KP_B}$ in the first place.

preceding section—we have just as much reason to doubt a form of DEC that ranges over Cartesian statements as we did to doubt unrestricted DEC, since none of the statements that gave us reason to doubt unrestricted DEC are such that knowing them is provably inconsistent. We doubt whether we can know them (or their negations), but we can't prove that we can't know them—they are by their nature Cartesian.

So far, so good. The problem with Tennant's restriction, however, is that it leaves the anti-realist open to new versions of the knowability paradox. Whereas Dummett's restriction ruled out too many statements, Tennant's doesn't rule out enough.

Here we need to return to full AR. In Tennant's restricted version:

$$(\mathrm{AR_C})\ \phi \leftrightarrow \Delta K\phi, \quad \text{where } \phi \text{ is Cartesian}$$

In particular, having so far only appealed to the left–right direction of anti-realism's central thesis (to derive KP), we will now need to appeal to its right-left direction. Recall that in Sect. 2 I said that the kind of modality involved in anti-realism's epistemic characterization of truth will have to be different from any familiar notion of logical, conceptual, or metaphysical modality. This is why. It would not do to characterize a true statement as one such that there is a logically/conceptually/metaphysically possible world in which it is known. For statements that are contingently *false* could satisfy *that* condition. Unlike standard conceptions of possible knowledge, and like knowledge itself, anti-realist knowability is *factive*. Here is Tennant (2000: p. 829)[9]:

> 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.

And Wright (2001: p. 60), in more Dummettian terminology:

> 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.

We can think of $\Delta K\phi$ as saying, roughly, that given how things are with us now in the actual world, it would be humanly feasible for someone at some time to perform investigative procedures so as to come to know ϕ.

A lot more would need to be said to make the notion precise but the intuitive idea is clear enough for present purposes.[10] What matters here, in addition to factivity, is that the following closure principle looks eminently plausible for this kind of possibility:

$$(\mathrm{CL})\ \Delta K\phi \wedge \Box(K\phi \rightarrow K\psi) \rightarrow \Delta K\psi$$

CL says: If, in the relevant anti-realist sense, it is possible for someone to know ϕ, and if every logically/conceptually/metaphysically possible world in which someone

[9] Cf. Tennant (2002: p. 140).

[10] As I read Kant, there is a close connection between anti-realist knowability and Kant's conception of 'the possible progress of experience' (e.g. at A492-3/B521). For relevant discussion, see Milmed (1967), Allais (2015: p. 142ff.), Stephenson (2015a), and Gomes and Stephenson (2016).

knows ϕ is also a world in which someone knows ψ, then it had better, in the relevant anti-realist sense, be possible for someone to know ψ too. It is important to be clear that, unlike its more familiar counterparts, CL 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, Δ is quite different from $\Diamond$ and so, in particular, is not the dual of $\Box$. Nevertheless, the intuitive plausibility of the normal schema carries over to CL. For in what sense could it be possible to know ϕ if it were not likewise possible to know something the knowing of which is a necessary condition of knowing ϕ?

At this stage, there are several ways to proceed. A number of new paradoxes have been developed and most of them require $\mathrm{AR_C}$ and CL plus some additional principles. I will present just one of these new paradoxes.[11] It requires only the following additional principle:

$$(*)\ \Box(K(\phi\wedge(K\phi\rightarrow K\psi))\rightarrow K\psi)$$

As a matter of logical/conceptual/metaphysical necessity, if it is known both that ϕ and that knowing ϕ implies knowing ψ, then it is known that ψ. Principle (*) assumes no more than was appealed to in the first stage of the original paradox from Sect. 2. Briefly: knowledge distributes over conjunction and is factive, so $K(\phi\wedge(K\phi\rightarrow K\psi))$ entails $K\phi\wedge(K\phi\rightarrow K\psi)$, which gives us $K\psi$ by elementary reasoning.

Now let p and q be basic, contingent statements. Then each of the following four statements is Cartesian:

$$q,\ \neg q,\ p\wedge(Kp\rightarrow Kq),\ p\wedge(Kp\rightarrow K\neg q)$$

That is, where p and q are basic and contingent, none of these statements is such that knowing it is provably inconsistent. Note also that both of the above conjunctions follow trivially from $p\wedge\neg Kp$, for false antecedents make for true material conditionals. As before, we begin by assuming such a statement for reductio:

(1) $p\wedge\neg Kp$	assumption for reductio

This time, however, we cannot directly apply our restricted anti-realist principle $\mathrm{AR_C}$, since statements of this form are not Cartesian. Instead, we appeal to the aforementioned Cartesian consequences of (1) and run two exactly parallel chains of reasoning, one for q and one for $\neg q$. First:

(2) $p\wedge(Kp\rightarrow Kq)$	1
(3) $\Delta K(p\wedge(Kp\rightarrow Kq))$	2, $\mathrm{AR_C}$ left–right
(4) $\Box(K(p\wedge(Kp\rightarrow Kq))\rightarrow Kq)$	(*)
(5) ΔKq	3, 4, CL
(6) q	5, $\mathrm{AR_C}$ right-left

[11] Due to Brogaard and Salerno (2006), building on work in Williamson (1992), Brogaard and Salerno (2002), and Rosenkranz (2004).

Second:

(2‘) $p \wedge (Kp \rightarrow K\neg q)$	1
(3‘) $\Delta K(p \wedge (Kp \rightarrow K\neg q))$	2‘, AR_{C} left–right
(4‘) $\Box(K(p \wedge (Kp \rightarrow K\neg q)) \rightarrow K\neg q)$	(*)
(5‘) $\Delta K\neg q$	3‘, 4‘, CL
(6‘) $\neg q$	5‘, AR_{C} right-left

We have our contradiction and the rest is as before:

(7) $\neg(p \wedge \neg Kp)$	1, 6, 6‘, discharging 1
(8) $p \rightarrow Kp$	7, classical logic
(9) $\forall p\ (p \rightarrow Kp)$	8, universal generalisation

Tennant’s restriction strategy is in trouble. Unlike our original omniscience claim from Sect. 2, the quantifier in (9) only ranges over basic, contingent statements. Still, that all such statements are known if true is hardly a palatable result for the anti-realist.

Again, this objection is not conclusive. Tennant (2009) has responded to this and other new knowability paradoxes by proposing further restrictions, some independent and some extensions or refinements of his Cartesian restriction. But the salient point here is just that Tennant’s restriction strategy looks less and less attractive with each reactionary addition. Two worries in particular are worth emphasizing. First, what’s to stop further paradoxes being developed that get around his specific, tailor-made restrictions? Second, the job of arguing that Tennant’s restriction strategy is principled and not ad hoc will be getting harder and harder with each such additional restriction. As before, we have motivation enough to search for an alternative restriction strategy that fairs better in these respects.

3.3 Non-epistemic statements

In the remainder of this paper I will defend a restriction strategy based on the following principle:

$$(\text{AR}_{\text{non-E}})\ \phi \leftrightarrow \Delta K\phi, \quad \text{where } \phi \text{ is non-epistemic}$$

A statement is non-epistemic when it makes no reference to the kind of cognitive capacities in terms of which anti-realism offers its epistemic characterization of truth. I expand on this below, but to a first approximation, we can think of non-epistemic statements as those that are K-free.[12]

[12] I should note that Tennant does briefly suggest (but then immediately rejects) a similar restriction for the right-left direction of AR. Since he does not propose the same for the left–right direction, he must retain his other restrictions, and since he conceives of ‘non-epistemic’ as simply K-free, he faces the belief problem I articulate below. Nor does Tennant address the question of principle for this restriction, instead quickly dismissing it as ‘rather drastic as a proposed logical inoculation’ (2009: p. 232). These last points

It is easy to see that this restriction yields a knowability principle strong enough to form the basis of the kind of Basic Revisionary Argument against classical logic that was sketched in Sect. 3.1. Combining a knowability principle restricted to non-epistemic statements with the classical law of excluded middle (unrestricted LEM) would yield a decidability theorem that ranges over non-epistemic statements:

$$(\mathrm{DEC}_{\text{non-E}})\ \Diamond K\phi \vee \Diamond K\neg\phi, \quad \text{where } \phi \text{ is non-epistemic}$$

And we have just as much reason to doubt $\mathrm{DEC}_{\text{non-E}}$ as we did to doubt unrestricted DEC, since none of the statements that gave us reason to doubt unrestricted DEC make reference to the kind of cognitive capacities in terms of which anti-realism offers its epistemic characterization of truth—they are all K-free. (Recall Dummett's original examples: Goldbach's conjecture and the continuum hypothesis; 'A city will never be built on this spot'; and 'If Jones had encountered danger, he would have acted bravely'.)

Moreover, this restriction yields a knowability principle that cannot be applied to statements of the form $\phi\wedge\neg K\phi$, which are not K-free, so it avoids the original version of paradox from Sect. 2. For the same reason, it avoids the new version of the paradox given in Sect. 3.2, which involved applying anti-realism's epistemic characterization of truth to statements of the form $\phi\wedge(K\phi\rightarrow K\psi)$. And, to the best of my knowledge, the same holds for all other extant versions of the paradox, since they all involve applying anti-realism's epistemic characterization of truth to statements that are not K-free.[13]

Indeed, we have reason to be cautiously optimistic that this is no accident and that no future paradox will be developed on the basis of $\mathrm{AR}_{\text{non-E}}$. This is because it is natural to think of the knowability paradoxes as manifesting a kind of self-reference phenomenon—anti-realism gives an epistemic characterization of truth and then gets into trouble when it is applied to epistemic truths. This is something that Alonzo Church already observed when he first discovered the paradox, noting that it 'is strongly suggestive of the paradox of the liar and other [as he then thought of them] epistemological paradoxes' (in Salerno 2009: p. 17). Church goes on to suggest that a solution appealing to the ramified theory of types might be appropriate. $\mathrm{AR}_{\text{non-E}}$ achieves the same general result by different, more local means. It is beyond the scope of this paper to determine the extent to which the knowability paradoxes really do exhibit self-reference phenomena.[14] But if they do, the present restriction strategy will stand us in especially good stead.

Footnote 12 continued
are connected—Tennant's mistake is to think that what we want is a logical inoculation, rather than a robust and principled form of anti-realism from which its own immunity to paradox naturally flows.

[13] See the references in fns.1 and 11.

[14] See Linsky (2009).

Before moving on to the question of principle, and thus finally turning to Kant, I should explain why thinking of non-epistemic statements as those that are K-free is only a first approximation of my official restriction. On its own it would not suffice. For suppose that belief is necessary for knowledge. Then a knowability principle applied to K-free statements would still be enough to yield the result that all K-free truths are believed. The reasoning is parallel to that involved in the original knowability paradox. Where B is our belief operator, statements of the form $\phi \wedge \neg B\phi$ are unknowable if knowledge entails belief—knowledge distributes over conjunction and is factive, so any statement of the form $K(\phi \wedge \neg B\phi)$ entails some statement of the form $K\phi \wedge \neg B\phi$, which in turn entails a contradiction if $K\phi$ entails $B\phi$. When ϕ is K-free, statements of the form $\phi \wedge \neg B\phi$ are also K-free. So if all K-free truths are knowable, then no such statement is true, which is just to say all K-free truths are believed. This 'omnicredence' result would be as unpalatable to the anti-realist as omniscience.[15]

A possible response here would be to reintroduce one of the previous restrictions—$\phi \wedge \neg B\phi$ is neither basic nor Cartesian (if $K\phi$ entails $B\phi$). Or we could deny *tout court* that belief is necessary for knowledge. Instead what I want to suggest is that, *insofar* as belief really is necessary for knowledge, then it involves the same kind of cognitive capacities in terms of which anti-realism offers its epistemic characterization of truth. Given the *official* statement of my restriction strategy—to statements that make no reference to the kind of cognitive capacities in terms of which anti-realism offers its epistemic characterization of truth—this means that $\mathrm{AR}_{\text{non-E}}$ can only be applied to statements that are both K- *and* B-free, which blocks the above derivation of omnicredence.

The background for the Kantian version of this view will come out in the next section, including why transcendental epistemology motivates exactly this restriction and not just one to K-free statements specifically. But to elaborate briefly on the point at hand, since it involves issues that will not be relevant in the next section: The fundamental analysandum for the transcendental epistemologist is the human capacity for knowledge—our *Erkenntnisvermögen*. This is an essentially rational capacity. When our rational capacity for knowledge functions well, it produces knowledge, a holding for true on subjectively and objectively sufficient grounds.[16] This is the concept of knowledge in terms of which the anti-realist who is also a transcendental epistemologist characterizes truth. But our capacity for knowledge is a fallible capacity and sometimes it malfunctions to produce mere belief, which is then understood as a holding for true on subjectively sufficient but objectively insufficient grounds. Belief per se—i.e. belief that is not necessarily *mere* belief—is then understood as a holding for true on subjectively sufficient grounds. And it is belief *in this sense* that is (analytically) necessary for knowledge *in this sense*. But for the transcendental epistemologist, all such states are *essentially* conceived of as various products of our

[15] My thanks to Lee Walters for pressing this worry. Note that related worries might arise for other cognitive conditions on knowledge, such as representation or thought. The response that follows generalizes.

[16] See A820ff./B848ff.; *Jäsche Logic* (9:66ff.).

essentially rational human capacity for knowledge, and so fall under our restriction. Kant's is a capacity-for-knowledge-first epistemology.[17]

Anti-Kantian anti-realists might not be able to adopt this kind of response to the omnicredence problem. But nor will they want to adopt $AR_{non\text{-}E}$ in the first place, at least not on the grounds I give in the next section. My concern here is with the anti-realist who is also a transcendental epistemologist.

4 Transcendental epistemology

I said in Sect. 2 that it is a serious question for any restriction strategy whether it is principled and not just ad hoc. Discussion of this issue with regard to Dummett's and Tennant's restriction strategies could be waived because they faced more straightforward problems. We have seen that our new restriction strategy fairs better in the relevant respects—it is of Goldilocksean strength in the sense that it yields a principle that is strong enough to form the basis of the anti-realist attack on classical logic but weak enough to avoid all extant (and, we can reasonably hope, future) versions of the paradox. So now we must face the question of principle. What grounds could the anti-realist have for restricting their epistemic characterization of truth to non-epistemic statements and so adopting $AR_{non\text{-}E}$?

To answer this question, I proceed as follows. First I consider what motivates anti-realism in the first place and refine our understanding of the view (Sect. 4.1). Then I outline a toy realist model that is meant to meet the anti-realist on their own terms (Sect. 4.2). Ultimately the model fails to force a realist concession from the anti-realist, but it is instructive because of its structure and its problems. Finally I introduce my own realist model for epistemic statements by appealing to Kant's doctrine of transcendental apperception (Sect. 4.3). I argue that the model meets the anti-realist on their own terms and that it doesn't suffer the problems of the previous model. Nor does it generalize to non-epistemic statements. This provides the anti-realist with a principled way to adopt $AR_{non\text{-}E}$: adopt transcendental epistemology and so concede a strictly limited realism for epistemic statements while retaining anti-realism for non-epistemic statements.[18]

[17] Note that none of this is to say that we cannot articulate an entirely naturalistic conception of belief, as a disposition to bet, say. And we might then think of belief in this sense as necessary for a kind of knowledge that we also understand in an entirely naturalistic way, as a true belief formed by a reliable mechanistic process, say. Beings that lack our essentially rational capacity for knowledge might enjoy states of this kind. As too might humans. And in humans, states of this kind might even be strongly correlated with states of the rational kind (A824-5/B853-4). But no such correlation is strictly necessary. So even though such naturalistic states might *not* fall under our proposed restriction, the derivations of omniscience and omnicredence from a *Kantian* anti-realism are still blocked.

[18] A note on the extent of my appeal to Kant in the following. I have argued elsewhere that Kant himself holds a form of anti-realism for empirical statements about appearances (Stephenson 2015a). I also think that Kant holds a form of realism for statements about things in themselves, with the broader view being that Kant is an anti-realist about all and only those statements about objects given to us through sensibility. I neither argue for nor rely on any of this here, however. For one thing, at least on the face of it, the motivations for Kant's anti-realism are quite different to those outlined in Sect. 4.1. What's important here is just that anti-realism about non-epistemic statements can be made compatible with realism about epistemic statements, which is what I argue for by appealing to Kant's account of apperception as providing

4.1 Anti-Realism and recognition-transcendence

Recall from Sect. 2 that realism and anti-realism agree that the meaning of a declarative statement is given by its truth-conditions but that realism places no epistemic constraints on truth so that a statement can be true independently of whether or not someone could, even in principle, come to know it. Realism thereby allows for statements whose meaning is given by truth-conditions that are *recognition-transcendent* in the sense that we might not, even in principle, be able to know whether or not they obtain. That is, realism allows for statements that instantiate the following schema:

$$\text{(RT)}\ \ (\phi \wedge \neg\Delta K\phi) \vee (\neg\phi \wedge \neg\Delta K\neg\phi)$$

It is the purpose of anti-realism's epistemic constraints on truth to rule out such statements—AR-type principles are incompatible with RT-type statements (modulo any corresponding restrictions). The canonical motivation for anti-realism and AR-type principles, then, comes from a pair of challenges to this realist conception of recognition-transcendence.

In a nutshell, suppose that we understand a statement whose meaning is given by recognition-transcendent truth-conditions. To understand a statement is to know what it means, so what we would have here is knowledge of the statement's recognition-transcendent truth-conditions. But how are we supposed to acquire or manifest knowledge of something that transcends our possible knowledge in this way? These are Dummett's acquisition and manifestation challenges to realism.[19] A little more fully:

We acquire knowledge of the meaning of a statement by learning how to use it, and we do this by learning to accept it as true in certain circumstances and reject it as false in others. This process can only involve conditions we can recognize as obtaining or failing to obtain. Recognition-transcendent conditions, by their very nature, can have played no part in such a process. How, then, can they form part of what we come to know when we come to know the meaning of a statement by learning how to use it?

Moreover, when we know what a statement means, we must be able to manifest that knowledge. Sometimes we can do so by giving an explicit, informative description of what the statement means using other words—'The cat is on the mat' means the feline is on the floor-covering. But on pain of regress, this cannot in general be the case. And an uninformative description—'The cat is on the mat' means the cat is on the mat—will not do because we can give these even when we have no idea what a statement means (or indeed when a statement is meaningless). In general, then, our knowledge of what a statement means will be implicit. It will consist in the possession of certain practical abilities that manifest in our use of the statement. When the meaning of a statement is given by truth-conditions that we can recognize as obtaining or failing to obtain,

Footnote 18 continued
a realist model of epistemic discourse that does not generalize. My only concern in this paper is to use this aspect of Kant to solve a problem for contemporary anti-realists.

[19] Dummett develops these challenges throughout his writings. See especially Dummett (1978: pp. 1–28, 215–247), Dummett (1981: p. 466ff.), and Dummett (1993: pp. 35–93). For useful discussion, see Wright (1993: pp. 13–23, 239–261), Hale (1997), Miller (2002), Murzi (2012), and Rumfitt (2015: p. 125ff.).

our implicit knowledge of its meaning will be manifest in our practical ability to discriminate between circumstances in which the statement is true and circumstances in which it is false. By the very nature of the case, we have no such ability when the meaning of a statement is given by recognition-transcendent truth-conditions. So in what practical ability could our knowledge of such a meaning be manifest?

Where no answer to these questions about acquisition and manifestation is forthcoming, the anti-realist infers that there can be no such thing as our understanding a statement whose meaning is given by recognition-transcendent truth-conditions. But then, the anti-realist continues, there can be no place in a theory of meaning for the notion of recognition-transcendent truth-conditions, since the point of a theory of meaning is to give an account of what we understand when we understand a statement. Whence the need for epistemic constraints on truth, and thereby meaning, embodied in some AR-type principle, which rules out the problematic conditions.

Looking at the motivation for anti-realism in this way helps bring out an important feature of the view that I have so far been able to gloss over, as it was not relevant to the issues so far discussed. The feature will be crucial for what follows, however. It is that the possible knowers in AR-type principles—the subjects whose possible knowledge that ϕ is equivalent to ϕ—must be *every understander* of ϕ (or at least some 'finite extension' of them, as Tennant puts it—see Sect. 3.2). What I mean is this. Suppose that you and I both understand ϕ but that our theory of what ϕ means allows that only you could possibly know that ϕ. This wouldn't be enough to satisfy the anti-realist. For the anti-realist, such a theory would leave it mysterious how I could possibly acquire or manifest the knowledge in which my understanding of ϕ supposedly consists, which is unacceptable.

To be clear, then, the anti-realist's acquisition challenge asks how *anyone* who knows what ϕ means could have acquired such knowledge if the meaning of ϕ is given by truth-conditions that *they* couldn't possibly recognize as obtaining or failing to obtain. The anti-realist's manifestation challenge asks how *anyone* who knows what ϕ means could manifest such knowledge if the meaning of ϕ is given by truth-conditions that *they* couldn't possibly recognize as obtaining or failing to obtain. Where no answer is forthcoming, the anti-realist places epistemic constraints on truth, and thereby meaning, which tie what *each subject* understands—i.e. the truth-conditions of ϕ—to what *they* could know—i.e. whether or not those conditions obtain. *This* is the kind of constraint required by anti-realism.[20]

Now, there are a number of ways realists might respond to these challenges, and thus resist anti-realism wholesale. They might object to the premises on which the challenges are constructed. Is meaning so closely connected to use? Must we be able to manifest our knowledge of meaning? Is the anti-realist right about the point of a theory of meaning? My aim here is not to mount a full defence of anti-realism, on these grounds or others, and I shall simply assume in what follows that the acquisition and

[20] With the quantification explicit, then, we have: (AR') $\forall\phi\forall s(sU\phi \rightarrow (\phi \leftrightarrow \Delta\exists t(stK\phi))$, where ϕ ranges over statements (restricted as appropriate), s over subjects (and finite extensions thereof), t over times, and where U says 'understands', K says 'knows', and Δ is our anti-realist possibility operator. This does not substantially affect any of the issues so far discussed, and since for what follows I only require the basic point that the possible knower has to be the understander, I won't update from the simpler formalization.

manifestation challenges are in order—I assume that, where these challenges cannot be met, anti-realism is warranted. My aim rather has been to provide anti-realism with a response to the particular problem posed for it by the knowability paradox, with the task now to show that the anti-realist has independent, philosophically robust motivation to concede my proposed restriction on their core principle.

To this end, what I want to do is show how Kant's transcendental epistemology provides us with the resources to *meet* the anti-realist's acquisition and manifestation challenges for epistemic statements, and in a way that would leave those challenges untouched for non-epistemic statements. The idea is that this justifies the proposed restriction of anti-realism's epistemic characterization of truth to non-epistemic statements, embodied in $\mathrm{AR}_{\text{non-E}}$. I will argue that transcendental epistemology enables the anti-realist to concede a strictly limited degree of realism about epistemic statements while maintaining anti-realism about non-epistemic statements.

By way of setting the stage for the Kantian motivation behind this realist restriction on anti-realism, let us first look at a related proposal due to Peter Strawson. It will provide a useful contrast case for my own proposal.

4.2 Pain and private ostension

Strawson (1977) suggests that the ascription of sensations to others constitutes a realist domain of discourse. Of course restricting anti-realism to statements that aren't about sensations wouldn't help much when it comes to the knowability paradox—knowledge isn't a sensation and statements of the form $\phi \wedge \neg K\phi$ (etc.) needn't be about sensations. But that's not the point of presenting the proposal. What's relevant is its structure and the problems it faces.

Here is Dummett's description of the proposal:

> On Strawson's view, I know what 'pain' means from my own case: when, so far as they could tell from the outward signs, I was in pain, others gave me the word, telling me, 'You are in pain'; but it is I who then invested the word with the meaning that it henceforth had in my language by means of a private ostensive definition, saying to myself, 'It is *this* that the word "pain" stands for'. Knowing, thus, from my own case what 'pain' means, I could now ascribe pains to others, even though I could in principle have no access to that which renders such ascriptions correct or incorrect. (1978: p. xxxii)

Dummett accepts that this would be a realist account of pain discourse. I know what 'Anil is in pain' means—'pain' refers to *this*, so 'Anil is in pain' means that things are with Anil as they are with me when I feel *this*. But unless I am Anil, there will be a gap between my knowledge of such meaning-constituting truth-conditions and my ability to know, even in principle, whether or not they are satisfied. For what determines whether or not they are satisfied, namely how things are with Anil, is something that I am not *in principle* able to access.

To be clear, none of this is to say, absurdly, that I *can't ever* know whether or not Anil is in pain. Of course I often can. But I must do so on the basis of Anil's behaviour, and this is what gives rise to the characteristic realist gap. For Anil's behaviour is only contingently related to his pain—'Anil is in pain' does not *mean* he is behaving in a certain way. He might be immobilized or feigning, and if he is, I might be unable, even in principle, to know whether or not he is in pain.

Yet I would still know what it means for him to be in pain—I have acquired this knowledge through private ostension and it is manifest in my practical ability to engage in public pain-talk as well as anyone. There is therefore no in principle connection between my grasp of the meaning of the statement and my ability to know whether or not it is true—the truth-conditions that constitute the meaning of my ascriptions of pain to others are potentially recognition-transcendent. This is a realist picture on which an AR-type principle that ranges over pain discourse would fail, since there could be an RT-type statement within that range. Presumably the account generalises to other sensations.

Or so the story goes. Unsurprisingly, Dummett rejects Strawson's proposal on the grounds that it 'unblushingly rejects that whole polemic of Wittgenstein's that has come to be known as "the private-language argument"' (1978: p. xxxii). Dummett focuses on what he sees as the incoherence of private ostensive definition. To this we can add the further, related worry that, even if private ostensive definition were *internally* coherent, so that I could come to know my own mind in this way, it immediately raises the conceptual problem of *other* minds. That is, even if we could give a word meaning through an act of private ostensive definition, it is far from clear that doing so would enable us to meaningfully apply that word in describing others. Two well-worn passages from Wittgenstein (1953) are often read as pressing this point:

> §283. What gives us *so much as the thought*: that beings, things, could feel something? Is it that my education has led me to it by drawing my attention to feelings in myself, and now I transfer the idea to objects outside myself?
>
> §302. If one has to imagine someone else's pain on the model of one's own, this is none too easy a thing to do: for I have to imagine pain which I *do not feel* on the model of the pain which I *do feel*

As I understand it, part of the issue here is that private ostension of my own pain provides no basis for the kind of distinction between the pain's *being* and its *being felt by me* that would be required of a general concept of pain, applicable not only to myself but to others as well. For pain presents subjectively (for private ostension) as a *mere modification* of my consciousness.[21] Thus the only concept I could possibly acquire in this way would be essentially indexed to me—it would not be the concept <pain> but rather the concept <*my* pain>. This is not a concept that it even makes sense to apply to others, for it makes no sense to think of *them* as feeling *my* pain.[22]

As with the acquisition and manifestation challenges themselves, I just want to grant that these are serious problems for Strawson's proposal so that it fails to provide

[21] As Kant puts it, sensation 'refers solely to the subject as a modification of its state' (A320/B376).

[22] See Bilgrami (1994) and Gomes (2011) for relevant discussion.

sufficient motivation for the anti-realist to concede realism about pain discourse (or sensation discourse generally). My proposal is that the transcendental epistemologist can provide a structurally similar realist model for epistemic discourse that meets the anti-realist's challenges while avoiding these problems. This provides the required motivation, *by the anti-realist's own lights*, for the kind of restriction of anti-realism to non-epistemic statements that is embodied in $AR_{non\text{-}E}$.

4.3 Rational activity and apperception

In parody of Dummett's parody, here is the basic story:

> On **Kant's** view, I know what '**know**' means from my own case: when, so far as they could tell from the outward signs, I **knew**, others gave me the word, telling me, 'You **know**'; but it is I who then invested the word with the meaning that it henceforth had in my language by means of **transcendental apperception**, saying to myself, 'It is **acting *like so*** that the word "**know**" stands for'. Knowing, thus, from my own case what '**know**' means, I could now ascribe **knowledge** to others, even though I could in principle have no access to that which renders such ascriptions correct or incorrect.

Why is this story any less problematic than Strawson's original? The key is that transcendental apperception of rational activity is very different from the kind of inner observation of pain to which Strawson appeals. Strawson's story was problematic in part because of its reliance on a strongly empiricist model of self-knowledge and in part because of the subjective nature of sensation. I will argue that my Kantian story does better in part because it develops a (moderately) rationalist model of self-knowledge and in part because of the objective nature of rational activity. In particular, and Dummett's Wittgensteinian worries notwithstanding, I will argue that apperception can provide us with general epistemic concepts that it makes sense to apply to others, even though in doing so, we apply them both beyond the conditions under which we acquired them and beyond the conditions under which we can know, as a matter of principle, whether or not they in fact apply.

First, some background. Kant's transcendental epistemology is concerned to analyse the human capacity for knowledge—our *Erkenntnisvermögen*. One of the central features of this analysis is the discernment, within the human capacity for knowledge, of two irreducibly different but intimately interconnected sub-capacities: a passive capacity for receptivity through the senses, called 'sensibility'; and an active capacity for spontaneity through concepts, judgement, and reason, called 'the understanding' (A50-2/B74-6).

It is its constitutive dependence on the understanding that makes the human capacity for knowledge an essentially rational capacity, more on which in a moment. It is its constitutive dependence on sensibility that makes the product of (successful) exercises of the human capacity for knowledge a kind of 'receptive' knowledge—it is knowledge of things that are in some way independent of or distinct from the particular act of knowing itself, information about which must be given to the knower through the senses (A19/B33). Kant sometimes calls receptive knowledge 'experience'. His

concern with the conditions for the possibility of experience is a concern with the conditions for the possibility of receptive knowledge.[23]

It is receptive knowledge that our Kantian story concerns. To repeat the first line of that story: On Kant's view, I know what 'know' means from my own case. Our first question, then, is how, according to Kant, do I know what 'know' in the receptive sense means from my own case?

Crucially, Kant's answer is *not* that I *receptively* know what receptive knowledge is from my own case. This would be to know through inner sense what receptive knowledge is, the Kantian correlate of an act of private ostensive definition—to acquire my concept of receptive knowledge by sensibly observing myself receptively knowing, from, as it were, outside that act of knowing. Such a model would likely face the same Wittgensteinian worries as Strawson's story about pain and so fail to motivate the anti-realist to concede realism about epistemic discourse.

Instead, for Kant, I *reflectively* know what receptive knowledge is from my own case. Receptive knowledge is a product of a rational capacity, and the key claim here is that exercising such a capacity *constitutively* involves *reflective* knowledge of the nature of *what one is thereby doing*, namely being actively *responsive to reasons* and judging (or acting) *for reasons*. Without such reflective knowledge, according to Kant, I simply would not be doing what I am doing in exercising a *rational* capacity. Hence my reflective knowledge of receptive knowledge, unlike my receptive knowledge itself, is not knowledge of something independent of or distinct from what is known. Reflective knowledge is rather knowledge that is partly *constitutive* of what is known—it is knowledge of what receptive knowledge is, from, as it were, within the act of receptively knowing. Our reflective knowledge of receptive knowledge is knowledge of the form, not the matter of receptive knowledge—it is knowledge through apperception, not the senses.[24]

Kant puts the distinction between inner sense and apperception in the *Anthropology* as follows:

> Inner sense is not pure apperception, a consciousness of what the human being *does*, since this belongs to the capacity for thinking. Rather it is a consciousness of what he *undergoes*, insofar as he is affected by the play of his own thoughts. (7:161; cf. B152-5, B157-9)

[23] E.g. at B147, B165–6, B218, B234, B277; *Prolegomena* (4:302). I do not mean to take a stance here on whether 'knowledge' rather than 'cognition' is a better translation of '*Erkenntnis*'. My claim is that the production of (receptive) knowledge is the primary function of our *Erkenntnisvermögen*, and I also take this kind of state to be included in (though not identical to) what Kant refers to as '*Wissen*' (see §2.3). These claims are compatible with allowing that the *Erkenntnisvermögen* can produce something that falls short of knowledge, yet which might still count as *Erkenntnis*. The capacity is essentially fallible (see §2.3). It can malfunction to produce states that are not justified or 'objectively sufficient', or which otherwise fail to 'agree' with their objects in the right way for knowledge. See Engstrom (2013: p. 39n.2) for this kind of view, and for further discussion of the general topic, see Gomes and Stephenson (2016), Willaschek and Watkins (2017), and Schafer (forthcoming). I myself have argued that hallucinations (which are not states of knowledge) count as *Erkenntnisse* in Stephenson (2015b) and Stephenson (2017).

[24] See especially Rödl (2007), Boyle (2009), Boyle (2011), Kitcher (2011), Kitcher (2017), and Leech (2017). For closely related discussion, see Smit (1999), Engstrom (2013), and Schafer (ms.).

Apperception is 'a consciousness of what the human being *does*'. As Kant describes it in the *Critique*:

> The consciousness of myself in the representation **I** is no intuition at all, but a merely **intellectual** representation of the *self-activity* of a thinking subject. (B278, my italics; cf. B278, B413)

The self-activity in question—what the human being *does*—consists in exercising her active, spontaneous, *rational* capacity, the understanding. It is because of his *constitutive*, *reflective* self-knowledge requirement on such activity that Kant calls the principle of apperception 'the supreme principle of all use of the understanding' (B136).

Now, as I understand Kant's theory of apperception, my reflective self-knowledge of what I am doing in receptively knowing needn't be total. When I receptively know that ϕ, I needn't reflectively know that I receptively know that ϕ. For one thing, I might be mistaken about *which* ϕ I receptively know. For another, I might be mistaken about whether I receptively *know* that ϕ, rather than merely believe that ϕ. There is no KK principle here. Nor must my reflective knowledge be explicit in the sense that I needn't be ready to fully articulate it in Kantian or any other jargon. But I do need at least implicit knowledge of the basic *rational* nature of my own activity in receptively knowing.

The preceding points forestall some immediate objections, but why countenance Kant's claim in the first place, that exercising a rational capacity constitutively involves reflective knowledge of the basic rational nature of what one is thereby doing in being actively responsive to reasons and judging (or acting) for reasons? This claim goes to the heart of the Critical philosophy. It shows up in the theoretical philosophy in Kant's account of the role of apperception in the rule-governed acts of synthesis that produce higher-order representations, including receptive knowledge.[25] It also shows up in the practical philosophy in the connections Kant draws between reason and autonomy.[26] It is not a claim I can fully defend here and there are different ways of doing so that yield different versions, and different strengths, of the claim.[27] But here is a way of putting the basic thought that suggests how congenial it might be to anti-realism generally, bearing in mind the origin of that view in constructivist mathematics (see Sect. 3.1). In exercising my rational capacity—as I do when I receptively know—I am actively *making up my mind*. And the kind of reflective self-knowledge through apperception that I have of this activity is a kind of maker's knowledge: it is knowledge of the nature of an activity that is had through engaging in and guiding that activity.

[25] See especially the Transcendental Deductions, A84ff. and B116ff. For my preferred account, see Evans, Sergot, and Stephenson (ms.).

[26] See especially the claim that a rational will can only act '*under the idea of freedom*', which is to say, it must represent itself, not as perfectly free or rational, but as at least *able* to act freely and thus for reasons, as not *inevitably* determined in its action by mere 'impulse' or 'alien influence' (*Groundwork* 4:448). For elaboration and defence, see Wood (2008: p. 130ff.).

[27] See the references in fn.24.

How, then, is this account of apperception relevant for our realist story of epistemic discourse, and thus for our proposed solution to the knowability paradox of restricting anti-realism to non-epistemic discourse by adopting $AR_{non\text{-}E}$?

The initial point is that, for Kant, I come to know what receptive knowledge is *through exercising* my capacity for such knowledge. More generally, it is in this way that I acquire my concepts of the products of rational capacities (which are conceived of as such), be they knowledge, belief (see Sect. 3.3), or something else (judgment, thought etc.). This needn't be an all or nothing affair. My rational capacity for receptive knowledge is innate, but I must learn how to exercise it, and I might do so gradually (*Jäsche Logic* 9:11). In particular, I must learn how to exercise its active, spontaneous part (its passive, receptive part takes care of itself). But again, this is something I learn how to do by doing—'it is a special talent that cannot be taught but only practiced' (A133/B172).[28] And as I gradually learn how to exercise my rational capacity for receptive knowledge, I *thereby* gradually come to know what such activity consists in and what it produces. *This* is how I know what 'know' in the receptive sense means 'from my own case', to refer back to the first line of our realist Kantian story again—through the very act of receptively knowing.

More needs to be said about the apperceptive process of acquiring reflective self-knowledge of the nature of our own rational activity in doing things like receptively knowing (believing etc.). But we need to be careful not to reify this process as its own, distinct activity.[29] If Kant is right that this kind of reflective self-knowledge is partly *constitutive* of rational activity—that possessing it is just part of what it is involved in, for example, receptively knowing—then an account of how we learn to receptively know *will already be* an account of how we acquire our reflective self-knowledge of what receptive knowledge (etc.) is.

Two further points beyond this acquisition claim are then required for my proposed application in our realist model of epistemic discourse. First, as noted above, this reflective knowledge needn't be explicit or theoretical knowledge. But if Kant is right, it is knowledge that is *manifest in my practical ability* to do things like receptively know, to believe, and to think and judge. Second, what I acquire and manifest through exercising my rational capacity are genuine, *general* concepts—my concepts of rational activity and of its products are concepts it makes sense to apply to others. This is possible because of how I have acquired these concepts—not through inner observation but through learning how to reason. For if Kant is right about rational activity and apperception, learning how to reason constitutively involves learning what *it is* to reason. Patricia Kitcher (2017: p. 170) puts both of these points very well:

> what subjects come to understand through engaging in higher cognition is not just how they apply concepts or make inferences, but how higher cognition

[28] Cf. *Anthropology* (7:199). Kant's reason for this claim is that we would be learning how to follow rules by following rules—see Ginsborg (2011) and Evans, Sergot, and Stephenson (ms.) for relevant discussion.

[29] Rödl (2007: p. 145) is especially clear on this. I suspect that several, otherwise excellent accounts of Kantian apperception and reflection are in danger of violating this proscription, e.g. Smit (1999), Westphal (2003), de Boer (2010), and Marshall (2014), though I cannot argue for this here, and it may well be that Kant himself either violates the proscription or at least uses the relevant terms to range over several different kinds of activity.

> works, and hence how any cognizer *must* think. They do not have a theory of thinking—they have no idea how these activities are possible. Rather, they have a practical understanding of what they do when they think. They apply that understanding to others and thus take everything that thinks to do what they do when they think.

The second point—about the generality of our apperceptively acquired concepts of rational activity—requires elaboration. Why does this Kantian account fair any better in the face of Dummett's Wittgensteinian worries than did Strawson's story about pain? The basic point is simply that the account does not appeal to anything like private ostension. Thus worries about either the incoherence of private ostension or the non-generality of concepts acquired through private ostension simply do not arise. But we can also say something much stronger and, for the anti-realist moved by such Wittgensteinian worries, more satisfying.

First, not only does the Kantian model not rely on private ostension; it positively *rules out* private ostension as so much as a *possible* route to concepts of rational activity. For the possibility of learning what rational activity is through private ostension of my own rational activity presupposes that I could perform such activity without already knowing what rational activity is—the picture would be that I perform rational activity, watch myself doing so, and only subsequently learn what rational activity is. Kant's constitutive, reflective self-knowledge requirement on rational activity rejects exactly this kind of division of labour.

Second, consider the nature of what I reflectively know on the Kantian model, of the kind of thing of which I have apperceptively acquired concepts. For Kant, a *rational* capacity is precisely a capacity to abstract from the peculiarities of my own situation, to pull myself free of mere impulse or alien influence and let myself be guided by *general* norms—to be rational in one's action just is to *universalize* the maxim for one's action (*Groundwork* 4:402ff.). As he puts it in the *Critique of the Power of Judgment*, we have in our rational capacity:

> a capacity for judging that in its reflection takes account (a priori) of everyone else's way of representing in thought, in order as it were to hold its judgment up to human reason as a whole… Now this happens by one holding his judgment up not so much to the actual as to the merely possible judgments of others, and putting himself into the position of everyone else, merely by abstracting from the limitations that contingently attach to our own judging (5:293–94)

Not only is the concept of rational activity that I acquire through the very exercise of my capacity for rational activity *not* essentially indexed to me; it is *essentially* not indexed to me—both the form *and the content* of the concept of rational activity is essentially general. This is quite different from the case of pain. Of course our actual concept of pain is indeed general—I understand my pain as the manifestation of a capacity for pain that could be shared by others. But the worry in the pain case was that, concerns about the internal coherence of private ostension aside, the most such a method of concept acquisition could get us would be a different, *non*-general concept of pain-*felt-by-me*. And the point here is that, once again, this isn't even a possibility on the present model. For there just is no non-general concept of rational

activity. Otherwise put, unlike the concept of pain-felt-by-me, the concept of rational-activity-performed-by-me is *already*, *necessarily* a concept of an activity that could be performed by someone else. For it is a concept of something I have done precisely by abstracting from the peculiarities of my own situation, by holding my judgement up to the possible judgements of others, to 'human reason as a whole'.

Here, finally, is the account in Kant's own words. The passage is from the Paralogisms chapter of the first *Critique*, where Kant's primary concern is to criticize the rational psychologists for claiming too much from apperception. But he begins by criticizing the empirical psychologists for starting from the wrong basis altogether, namely an inner sense model of self-knowledge:

> But right at the start it must seem strange that the condition under which I think in general, and which is therefore merely a property of myself as subject, is at the same time to be valid for everything that thinks, and that upon a statement that seems empirical we can presume to ground an apodeictic and universal judgment, namely: that everything that thinks is so constituted as the claim of self-consciousness asserts of me. The cause of this, however, lies in the fact that we must necessarily ascribe to things a priori all of the properties that constitute the conditions under which alone we think them. Now I cannot have the least representation of a thinking being through an external experience, but only through self-consciousness. Thus such objects are nothing further than the transference of this consciousness of mine to other things, which can be represented as thinking beings only in this way. (A346-7/B404-5)[30]

In the terms of our realist account of epistemic discourse: I (reflectively) know what 'Anil (receptively) knows' (or 'believes' or 'thinks' or 'judges' etc.) means. It means that things are with Anil as they are with me when I (receptively) know (etc.). For 'everything that thinks is so constituted as the claim of self-consciousness asserts of me…such objects are nothing further than the transference of this consciousness of mine to other things'. But unless I am Anil, there will be a gap between my understanding of such meaning-constituting truth-conditions and my ability to know, even in principle (and in any way), whether or not they are satisfied. For what determines whether or not they are satisfied, namely how things are with Anil, is something that I am not *in principle* able to access.

To be clear, none of this is to say, absurdly, that I *can't ever* know whether or not Anil knows. Of course I often can. But I must do so on the basis of Anil's behaviour, and this is what gives rise to the characteristic realist gap. For Anil's behaviour is only contingently related to his knowledge—'Anil knows' does not *mean* he is behaving in a certain way. He might be immobilized or feigning good reasoning though he just got lucky, and if he is, I might be unable, even in principle (and in any way), to know whether or not he knows.

Yet I would still (reflectively) know what it means for him to (receptively) know—I have acquired this (reflective) knowledge through transcendental apperception, and even if I cannot articulate it, perhaps because I have not read Kant, I can and do manifest

[30] For further relevant discussion of this passage, see Rödl (2007: p. 181ff.) and Engstrom (2013: p. 52f.).

that (reflective) knowledge in my practical ability to exercise my own rational capacity for (receptive) knowledge, for there can be no such (receptive) exercise without such (reflective) knowledge.

There is therefore no in principle connection between my grasp of the meaning of the statement 'Anil (receptively) knows' and my ability to know (in any way) whether or not it is true.

And note that the point generalizes from particular positive applications to general negative ones. If I might be unable even in principle to know whether or not Anil (receptively) knows, then, since Anil not (receptively) knowing is a condition on no-one (receptively) knowing, I might be unable even in principle to know when no-one (receptively) knows, thus unable even in principle to know statements of the form $\neg K\phi$.[31] I would still (reflectively) know what such statements mean—they mean that it is not the case with anyone that things are with them as they would be with me were I to (receptively) know ϕ.

More generally, then, the truth-conditions that constitute the meaning of my ascriptions of rational epistemic states (or lack thereof) to others are potentially recognition-transcendent. This is a realist account of epistemic discourse on which an AR-type principle *that ranges over epistemic statements* would fail, since there could be RT-type statements within that range.

However, the account does *not* generalise to *non*-epistemic statements. What I come to reflectively know through transcendental apperception of my own rational activity is the nature of rational activity as such, of what it is to be actively responsive to reasons and to judge (or act) for reasons, as I and you and others do when we do things like receptively know. This reflective knowledge is a kind of conceptual knowledge. It is knowledge of what it means to exercise a rational capacity and of concepts such as <receptive knowledge>, <belief>, <judgment>, and <thought>, where these are understood as concepts of the various products of rational activity. It is not knowledge of the concepts involved in *non*-epistemic statements. Transcendental epistemology provides no response to Dummett's acquisition and manifestation challenges for such statements. It won't help with the question of how I could acquire or manifest knowledge of the meaning of 'There are inhabitants of the moon' (A492/B521) or 'All bodies are heavy' (A7/B11), were that meaning supposed to be given by truth-conditions that I couldn't possibly recognize as obtaining or failing to obtain.

Here, finally, we have a *principled* way for the anti-realist to adopt $\mathrm{AR}_{\text{non-E}}$: they adopt transcendental epistemology and concede a strictly limited realism for epistemic statements while retaining their anti-realism for non-epistemic statements. The result would be a thoroughly anti-realist picture, one that gives an epistemic characterization of truth and meaning for a vast swathe of discourse, and yet which is not susceptible to the knowability paradox.

[31] Note that this does not rely on a *general* closure principle for knowledge, since all that is required to block knowability is that I *might* be unable to know.

5 Conclusion

The knowability paradox poses a serious problem for anti-realism by threatening to collapse the core principle of the view into an unacceptable omniscience claim. I have argued that Kant's transcendental epistemology provides anti-realism with the resources to solve this problem. The proposal was that we restrict anti-realism's epistemic characterization of truth to statements that make no reference to the kind of cognitive capacities in terms of which that characterization is given. The first stage in the argument was to show that this restriction strategy fairs better in certain quasi-formal respects than do other prominent restriction strategies (Sect. 3). The second stage in the argument was to show that the proposed restriction is philosophically principled (Sect. 4). It amounts to conceding realism about epistemic statements while maintaining anti-realism about non-epistemic statements. This is where I appealed to transcendental epistemology: to motivate such a compromise.

Dummett said in a valedictory lecture on realism and anti-realism: 'I viewed my proposal, and still continue to view it, as a research programme… as the posing of a question how far, and in what contexts, a certain generic line of argument could be pushed' (1993: p. 464). There is an echo here of the 'Copernican experiment' that Kant considered an 'altered method of our way of thinking' (Bxvi-xix). It proves the key to my proposed solution to the knowability paradox. For if Kant is right about apperception, I have argued, then although the anti-realist argument might be pushed very far indeed, it cannot be pushed so far that it collapses into omniscience. More needs to be said in elaboration and defence of transcendental epistemology, in particular its central claim that exercising a rational capacity constitutively involves reflective self-knowledge of the nature of what one is thereby doing. But the prospects for a transcendental anti-realism look good.[32]

References

Allais, L. (2015). *Manifest reality: Kant's idealism and his realism*. Oxford: Oxford University Press.

Bilgrami, A. (1994). Dummett, realism and other minds. In B. McGuinness (Ed.), *The philosophy of Michael Dummett* (pp. 205–228). Alphen aan den Rijn: Kluwer.

Boyle, M. (2009). Two kinds of self-knowledge. *Philosophy and Phenomenological Research, 78*(1), 133–164.

Boyle, M. (2011). Transparent self-knowledge. *Proceedings of the Aristotelian Society, 85,* 223–241.

Brogaard, B., & Salerno, J. (2002). Clues to the paradoxes of knowability: reply to Dummett and Tennant. *Analysis, 62*(2), 143–150.

Brogaard, B., & Salerno, J. (2006). Knowability and a modal closure principle. *American Philosophical Quarterly, 43*(3), 261–270.

[32] For extremely helpful comments on earlier drafts of this material, my thanks to Anil Gomes, Nora Kreft, Lee Walters, Jack Woods, and audiences in Berlin, Oxford, Southampton, and Vienna, as well as to two thorough and challenging anonymous referees for this journal.

Brogaard, B. & Salerno, J. (2013) 'Fitch's Paradox of Knowability', E. N. Zalta (ed.), *The Stanford encyclopedia of philosophy*: https://plato.stanford.edu/archives/win2013/entries/fitch-paradox/.
Brouwer, L. E. J. (1975) *Collected works I.* In A. Heyting (ed.), North-Holland.
de Boer, K. (2010). Pure reason's enlightenment: Transcendental reflection in Kant's first critique. *Kant Yearbook, 2,* 53–73.
Dummett, M. (1978). *Truth and other enigmas*. Harvard: Harvard University Press.
Dummett, M. (1981). *The interpretation of Frege's philosophy*. London: Duckworth.
Dummett, M. (1991). *The logical basis of metaphysics*. Harvard: Harvard University Press.
Dummett, M. (1993). *The seas of language*. Oxford: Oxford University Press.
Dummett, M. (2000). *Elements of intuitionism*. Oxford: Oxford University Press.
Dummett, M. (2001). Victor's error. *Analysis, 61*(1), 1–2.
Engstrom, S. (2013). Unity of apperception. *Studi Kantiani, 26,* 37–54.
Evans, R., Sergot, M., and Stephenson, A. (ms.) 'Formalizing Kant's Rules: a logic of conditional imperatives and permissives'.
Ginsborg, H. (2011). Primitive normativity and skepticism about rules. *Journal of Philosophy, 108*(5), 227–254.
Gomes, A. (2011). Is there a problem of other minds? *Proceedings of the Aristotelian Society, 111,* 353–373.
Gomes, A., & Stephenson, A. (2016). On the relation of intuition to cognition. In D. Schulting (Ed.), *Kantian nonconceptualism* (pp. 53–79). Palgrave-Macmillan: Basingstoke.
Hale, B. (1997) Realism and its oppositions. In B. Hale & C.Wright (eds.), *A companion to the philosophy of language*, Hoboken: Blackwell, pp. 271–308.
Hilbert, D. (1902). Mathematical problems. *Bulletin of the American Mathematical Society, 8,* 437–479.
Kant, I. (1902-) *Gesammelte Schriften*, Deutschen Akademie der Wissenschaft, de Gruyter.
Kitcher, P. (2011) *Kant's Thinker*, Oxford: Oxford University Press.
Kitcher, P. (2017). A Kantian critique of transparency. In: A. Gomes & A. Stephenson (eds.), *Kant and the philosophy of mind: Perception, reason, and the self* , Oxford: Oxford University Press, pp. 158–172.
Leech, J. (2017). Judging for reasons. In: A. Gomes & A. Stephenson (eds.), *Kant and the philosophy of mind: Perception, reason, and the self* , Oxford: Oxford University Press, pp. 173–188.
Linsky, B. (2009). Logical types in arguments about knowability and belief. In: J. Salerno (ed.), *New essays on the knowability paradox*, Oxford: Oxford University Press, pp. 163–179.
Marshall, C. (2014). Does Kant demand explanations for all synthetic a priori claims? *Journal of the History of Philosophy, 52*(3), 549–576.
Miller, A. (2002). What is the manifestation argument? *Pacific Philosophical Quarterly, 83,* 352–383.
Milmed, B. (1967). "Possible Experience" and reecent interpretations of Kant. *The Monist, 51*(3), 442–462.
Moore, A. W. (2012) *The evolution of modern metaphysics: Making sense of things*, Cambridge University Press.
Murzi, J. (2012). Manifestability and epistemic truth. *Topoi, 31,* 17–26.
Putnam, H. (1981). *Reason, truth and history*. Cambridge: Cambridge University Press.
Rödl, S. (2007). *Self-consciousness*. Harvard: Harvard University Press.
Rosenkranz, S. (2004). fitch back in action again? *Analysis, 64*(1), 67–71.
Rumfitt, I. (2015). *The boundary stones of thought*. Oxford: Oxford University Press.
Salerno, J. (ed.). (2009) *New Essays on the Knowability Paradox*, Oxford University Press.
Schafer, K. (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.
Smit, H. (1999). The Role of Reflection in Kant's *Critique of Pure Reason*. *Pacific Philosophical Quarterly, 80,* 203–223.
Stephenson, A. (2015a). Kant, the Paradox of Knowability, and the Meaning of 'Experience'. *Philosophers' Imprint, 15*(27), 1–19.
Stephenson, A. (2015b). Kant on the Object-Dependence of Intuition and Hallucination. *The Philosophical Quarterly, 65*(260), 486–508.
Stephenson, A. (2017). Imagination and Inner Intuition. In: *Kant and the philosophy of mind: Perception, reason, and the self* , eds. A. Gomes & A. Stephenson, Oxford: Oxford University Press, pp. 104–123.
Strawson, P. F. (1966) *The bounds of sense*, Abingdon: Routledge.
Strawson, P. F. (1977). Scruton and Wright on anti-realism etc. *Proceedings of the Aristotelian Society, 77,* 15–22.
Tennant, N. (1997) *The taming of the true*, Oxford: Oxford University Press.

Tennant, N. (2000). Anti-Realist Aporias. *Mind, 109*(436), 825–854.
Tennant, N. (2002). Victor Vanquished. *Analysis, 62,* 135–142.
Tennant, N. (2009) 'Revamping the restriction strategy', in Salerno (ed.), *New essays on the knowability paradox*, Oxford University Press, pp. 223–238.
Walker, R. (1995). Verificationism, anti-realism, and idealism. *European Journal of Philosophy, 3*(3), 257–272.
Westphal, K. (2003). Epistemic reflection and cognitive reference in Kant's transcendental response to scepticism. *Kant-Studien, 94,* 135–171.
Willaschek, M., & Watkins, E. (2017). Kant on cognition and knowledge. *Synthese*. https://doi.org/10.1007/s11229-017-1624-4.
Williamson, T. (1992). On intuitionistic modal epistemic logic. *Journal of Philosophical Logic, 21,* 63–89.
Williamson, T. (2009). 'Tennant's Troubles'. In Salerno (Ed.), *New Essays on the Knowability Paradox* (Vol. 4, pp. 183–204). Oxford University Press.
Wittgenstein, L. (1953) *Philosophical Investigations*, G. E. M. Anscombe (trans.), Macmillan.
Wood, A. (2008). *Kantian ethics*. Cambridge: Cambridge University Press.
Wright, C. (1992). *Truth and objectivity*. Harvard: Harvard University Press.
Wright, C. (1993). *Realism, meaning and truth*. Hoboken: Blackwell.
Wright, C. (2001). On being in a quandary. *Mind, 110,* 45–98.
Schafer, K. (ms.) 'Spontaneity, Self-Consciousness, and Maker's Knowledge in Kant'.