Epistemic Paradox and the Limits of Knowledge

James R. Shaw[†]

Abstract

Caie (2012) argues that a form of epistemic paradox drives paracomplete theorists of paradox to accept that we can be rationally required to have indeterminate beliefs. I defend a more moderate position on which the case requires only agnosticism. Part of the defense involves identifying a controversial, and undefended, tacit premise in Caie's reasoning that the moderate position is untenable. After doing so, I conclude by noting that even the moderate position I defend shares a striking consequence with Caie's view: omniscience becomes metaphysically impossible, an epistemic limitation which would complement that from Fitch's Paradox.

Caie (2012) uses a form of epistemic paradox to argue for the surprising claim that we can be rationally required to have indeterminate beliefs. I defend a more moderate response to the paradox, on which we are driven instead to agnosticism. In particular, I note that Caie's argument against such a position smuggles a tacit premise into a transparency principle. This tacit premise either begs the question against the moderate position, or posits the existence of highly controversial propositions with something like infinite nested structure. Though the moderate position escapes Caie's argument, I concede that it shares an unusual consequence with Caie's view: omniscience becomes metaphysically impossible—an epistemic restriction distinct from, and complementing, that given by Fitch's Paradox.

1 Epistemic Paradox and Evidence

Let "Ada" name some competent, reflective, English-speaking reader of this paper. Let *t* be the moment she finishes reading it, and consider (1).

Draft of March 2, 2016; please don't cite without permission
✉: jrs164@pitt.edu

† For helpful discussions of this material, I'm grateful to Michael Caie, Ulf Hlobil, Dustin Tucker.

(1) At *t*, Ada doesn't believe (1) is true.

Caie (2012) notes that sentences like (1) generate an epistemic paradox by rendering three plausible claims about rationality inconsistent with a contingent transparency assumption about Ada's ability to be aware of her own beliefs.[1]

> *Evidence*: If an agent's evidence entails *p*, she ought to believe *p* (on pain of irrationality).
>
> *Consistency*: It's irrational to believe both *p* and not *p*.
>
> *Possibility*: It's always possible to be rational.
>
> *Transparency*: If Ada believes *p*, she believes that she believes *p*, and it is among her evidence that she believes *p*. Likewise, if Ada does not believe *p*, she believes that she does not believe *p*, and this is among her evidence.

Suppose Ada believes (1) is true at *t*. The following is a truth of semantics: if Ada believes (1) is true at *t*, (1) isn't true. So two propositions—that Ada believes (1) true, and the semantic fact just mentioned—entail that (1) isn't true. But both of those propositions (in part by *Transparency*) seem to be among Ada's evidence at *t*. So if Ada doesn't also believe (1) *isn't* true at *t*, *Evidence* counts her irrational. And if instead she manages to believe (1) isn't true, while continuing to also believe it true, *Consistency* will count her irrational instead. So whatever she does, she can't rationally believe (1) is true at *t*.

Now suppose instead Ada doesn't believe (1) is true at *t*. Then, again as a semantic matter, (1) will be true. Again, since Ada is aware of the semantics of (1) at *t* and *Transparency* holds, *Evidence* counts her irrational (since, *ex hypothesi* she doesn't believe (1) is true, as *Evidence* would then require).

What all this seems to show is that if Ada really does satisfy *Transparency*, and her only options at *t* are to believe (1) true or not, then the first two theses above guarantee she is irrational at *t*, contradicting *Possibility* (at least on a plausible reading of that principle).

---

[1] The importance of cases like (1) was appreciated by Burge (1978, 1984), who notes the puzzle case goes back to Buridan. I've reworked Caie's particular argument a little, but not in a way that I think matters for my claims here.

Which principle is at fault? *Transparency* as I've stated it is a very strong assumption, for familiar reasons. For example, it may require Ada to have infinite hierarchies of beliefs. And it seems to claim, of any proposition Ada hasn't entertained, that she's aware she doesn't believe it. But Caie notes that all we really need to recreate epistemic paradox are metaphysically possible agents for whom *Transparency* holds. I agree that the relevant agents, or something close enough, are possible. Indeed, I suspect a modest enough form of transparency generating paradox would even be plausible for ordinary readers of this paper, like Ada, as regards beliefs about (1) at *t*. All we need is for Ada to be reliably aware of what she thinks about (1) at *t*, which certainly seems possible. So I don't think we should find fault there.

Some have endorsed the existence of 'rational dilemmas' that would violate *Possibility* (Gibbard & Harper (1978), Priest (2002), Ross (2010)). And dialetheists, like Priest (1985, 2006), are happy to embrace rational contradictory beliefs, jettisoning *Consistency*. But, with Caie and many others, I do not find these positions appealing. At the very least, I do not find them plausible as resolutions of this case.

What options remain? Caie puts forward an inventive proposal that safeguards all principles by abandoning bivalence for claims of belief. Here is the idea: Not only can Ada believe (1) true, or not, but she can be in a doxastic state that is indeterminate with respect to both. If so, this would block the final step of the argument above as involving a kind of false dilemma between belief and non-belief. Nor can we obviously reconstruct paradox once Ada is in this indeterminate state. Suppose at *t*, it is indeterminate whether Ada believes (1) true. Then even if she's aware of her indeterminate beliefs at *t*, this awareness won't conjoin with her semantic evidence about (1) to drive her into inconsistency. If it's among her evidence that she doesn't determinately believe, or determinately not believe, that (1) is true, what apparently follows is simply that it is indeterminate whether (1) is true. Thus the evidence she possesses at *t* does not seem to be grounds for adopting a full-on belief or disbelief in (1)'s truth. Arguably, it doesn't mandate agnosticism either (after all, agnosticism involves determinate lack of belief). If that's right, Ada can precariously, but rationally, teeter between commitments about (1). Indeed, on Caie's proposed resolution, Ada becomes rationally required at *t* to indeterminately believe (1) true. That is the only way she can stay rational.

Do these indeterminate doxastic states make sense? What are they like? One positive example Caie cites are vague beliefs, which I think we should agree are possible. It's not only slowly dwindling heaps of sand and gradually balding men that provide fodder for Sorites arguments. Between the clear applications of virtually any term and its negation we can construct gradual degrees of change creating borderline cases. Attributions of belief should be no different.

Consider Susie. If you ask Susie, she will sincerely, but half-heartedly, say broccoli tastes good. But most of the time she'll refuse to eat it, acting (sincerely) as if it's very unappetizing. Odd things change this behavior. If broccoli is served alongside fruit or starches she gobbles it up eagerly (even if she doesn't touch the other food). Also, she can only distinguish broccoli from zucchini about 65% of the time. Now: does Susie believe that broccoli tastes good or not? I think this is, or comes close, to indeterminacy—a borderline case of belief we can't settle either way.

Caie claims Ada is required, at *t*, to engage with (1) a bit like Susie engages with broccoli. She shouldn't determinately believe (1). But she shouldn't determinately not believe it either. She should aim to get into a place where it's a little tricky to tell what she believes.

I don't see anything incoherent with Caie's position, but I do think it may be under-motivated. I am suspicious of the claim that sentences like (1) rationally foist us into odd indeterminate doxastic states like Susie's. Pre-theoretically, I see no irrationality in Ada's seeing and understanding (1), then simply ignoring it at *t*. Theoretically, this drives me to a view prefigured in Conee (1987) and Richter (1990): abandoning *Evidence*.

As Caie notes, *Evidence* is a synchronic constraint.[2] But Caie does not discuss, as he does for issues of transparency, that as such *Evidence* is too strong to apply to agents like us. Inferences are costly in cognitive resources. Mathematicians aren't to be rationally faulted for not believing every consequence of the Peano Axioms—especially not the boring and inconsequential ones. Also inferences take time.[3] Suppose you believe by sight that it is now precisely

[2] *ibid.*, p.5, n.13.

[3] See, e.g., Williamson (2000) p.282 and, for the point applied the present puzzle, Conee (1987). The claim seems to be entailed by views that describe inference as a process like Boghossian (2014), Broome (2014), Wright (2014). (For a recent noteworthy detractor see Neta (2013).) It is worth flagging, however, that the sense in which I appeal to the claim that inference takes time here (i.e., time passes between acquiring a belief and drawing conclusions on the basis of it) needn't come into conflict with the claim of White (1971) that inference is not a process. Also relevant to this point is the exchange between Sorensen

10:45:32am, and have a standing belief that the bomb detonates at precisely that second. It follows from your *de se* and *de dicto* beliefs that the bomb is detonating in the current second. But the second it might take you to infer this is the second that will make that *de se* belief knowably false. How could you be rationally faulted for failing to acquire that knowably incorrect *de se* attitude?

It is tempting to claim, as we did for issues of transparency, that the obstacles here are irrelevant metaphysical contingencies that could be overcome by super-beings. But this is much less obvious than before. The first worry isn't necessarily allayed by granting us limitless cognitive resources. Even if we could infer without cost, are you sure we would be *irrational* for not believing every single odd, lengthy, foreseeably unusable logical entailment of our evidence?

More importantly, the second problem about time isn't just a problem about our cognitive limitations either, but about the metaphysics of inference. To overcome it, we need beings who can simultaneously come to a new belief and infer its consequences. And *that* may indeed be metaphysically impossible. It's not clear that a simultaneous acceptance of the consequences of a belief and the belief itself could count as having inferred the consequences from the premises. And, even if it does so count, it isn't obvious that such 'inferences' would grant their 'conclusions' the relevant relations of rational support for which good inference is prized. If either of these worries about timing is well founded, and we maintain *Evidence*, then in the instant anyone (including super-beings) acquire any new belief, they will be counted irrational—either by *Evidence*, for failing to believe consequences of their new belief at that moment, or by believing those consequences without proper rational support.

One might object that the above worries only require us to weaken *Evidence* in certain ways. Couldn't the resulting weakened principle recreate paradox? I grant that we may need a replacement principle, since the situation with *Evidence* seems in some ways analogous to that with *Transparency*. Most recognize that even though *Transparency* is too strong to apply to agents like us, we still have some kind of special access to our own mental states that calls out for explanation, perhaps in the form of some weakened or highly circumscribed principle.[4] Likewise *Evidence*, though too strong, seems to latch on to an important normative truth: there appear to be important rational requirements

(1987) (see p.312) and Richter (1990).

4 See Byrne (2005, 2011) for careful formulations of the explanatory need.

on how we expand our beliefs in light of evidential entailments.

But if the second of my two worries about *Evidence* is founded, relevantly weakened versions of the principle will probably have escape clauses to reconcile it with *Transparency*, *Consistency*, and *Possibility*. A suitably restricted version of *Evidence* might look very roughly like the following principle.

> *Evidence*$^-$: If (i) an agent's evidence entails p and (ii) p or that evidence aren't liable to be false once the inference is performed, then the agent ought to *come to* believe p on the basis of that evidence.

But note that we can no longer derive any problems from this along with the rest of Caie's principles. If Ada is agnostic, the truth of (1) still follows from facts about Ada's epistemic state at t, of which she'll be aware. But if, at t itself, Ada hasn't yet come to believe (1), *Evidence*$^-$ needn't count her irrational for so doing. If she didn't believe (1) just before t, inferring (1) would have rendered (1) false, or disrupted any justification she had for taking it to be true. The inference would be in violation of the second caveat. So *Evidence*$^-$ won't count her irrational for not making the inference. It's possible that at t itself *Evidence*$^-$ directs her to infer (1)'s truth. But it would only do this if, by the time the inference were done, it posed no threat to her rationality because t had passed.[5]

I suspect some such worries will give us grounds to reject *Evidence* in a way that rationally permits Ada's agnosticism about (1)'s truth at t. But I needn't rely on this suspicion. The strongest reason to reject *Evidence* comes from epistemic paradox itself. I recently noted the rough idea behind *Evidence*, with which I sympathize: that there are important rational constraints on how we should expand our beliefs in light of evidential entailments. We can be blamed for failing to derive obvious, foreseeably relevant consequences of our available evidence, especially if the epistemic costs of deriving those consequences are relatively low. But this justification evaporates when considered on the 'agnostic' horn of the epistemic dilemma involving (1). Suppose Ada ignores (1) at t—effectively remaining agnostic concerning its truth. Then, recall, *Evidence* pronounces Ada irrational at t for not believing (1) true, effectively on the basis of the evidence that she doesn't then believe it. But surely if Ada had done what

[5] The basic strategy here, as alluded to before, isn't novel. Again, see Conee (1987), Richter (1990). Note that this way out persists even if (1) isn't indexed to a particular time, but a span: be agnostic at as many problematic times in the span as necessary.

Evidence prescribed—believed (1) true—then the relevant evidence wouldn't have been around. Indeed, it is metaphysically impossible for Ada to believe (1) true at *t* on the basis of true evidence that she doesn't. What this shows is that the intuitive grounds for holding *Evidence* simply don't cover the case of (1). We should reject *Evidence* because of epistemic paradox itself. The reasoning in the puzzle reveals that, and why, the principle is too strong.

Caie anticipates these kinds of maneuvers, especially the last:

> I can imagine the following response seeming attractive...the above paradox shows...that in certain cases one can have evidence that makes it certain that a particular proposition is true, but in such a case one's having that evidence essentially depends on one's *not* responding to the evidence by believing the proposition in question. At least in such cases, according to this line of thought, evidence does not rationally mandate belief.[6]

But Caie claims such a response can't work and *Evidence* can't be the source of our problems. This is because we can generate epistemic paradox, bypassing the reliance on *Evidence*, if we recreate the puzzle with propositions instead of sentences.

Let me drop reference to time, as a slight idealization. Then consider (2).

(2) Ada doesn't believe the proposition expressed by (2).

We can argue as follows. First we note the following equality between sentences, true by definition.

(a) (2) = "Ada doesn't believe the proposition expressed by (2)"

Now suppose (2) for reductio.

(b) Ada doesn't believe the proposition expressed by (2).

By *Transparency*, Ada then believes: Ada doesn't believe the proposition expressed by (2). That is,

(c) Ada believes the proposition expressed by "Ada doesn't believe the proposition expressed by (2)".

[6] Caie (2012) p.11.

But by the equality in (a), we have

(d) Ada believes the proposition expressed by (2).

But (d) contradicts (b). This seems to show, by reductio, that it is impossible for Ada to simply not believe the proposition expressed by (2) while satisfying *Transparency*. If she believes that she doesn't believe the proposition expressed by (2), she *ipso facto* believes it. But now if she believes the proposition expressed by (2), as per (d), by *Transparency*

(e) Ada believes the proposition expressed by "Ada believes the proposition expressed by (2)".

But by (d) and (a) we have

(f) Ada believes the proposition expressed by "Ada doesn't believe the proposition expressed by (2)".

But (e) and (f) attribute contradictory beliefs to Ada. So again, she's doomed to inconsistent belief, violating *Possibility* on the assumption of *Consistency*. And the deduction seemingly made no appeal to *Evidence*.

If the foregoing argument were just as good as the non-propositional version of epistemic paradox, Caie would be right to ignore *Evidence* as irrelevant to the puzzle. But the foregoing argument is highly problematic, precisely because it attempts to bypass that premise. Let me first discuss a minor dialectical worry before getting to the real problems.

The dialectical worry is that Caie's argument relies on the tacit assumption that (2) expresses a proposition. This *could* be controversial among some of Caie's intended targets which include "anyone who advocates a paracomplete treatment of the semantic paradoxes"—that is, a treatment that denies excluded middle.[7] Some such theorists take liar-like paradoxical sentences to fail to express propositions altogether. Glanzberg (2001) is a recent example, but the view goes back to the first systematic paracomplete treatment of paradox in Kripke (1975).[8]

---

7 *ibid.* p.2.

8 See, in particular, *ibid.* pp.699ff., and especially p.700 n.18. Oddly, Caie says that we can implicitly see in Kripke's treatment of paradox the view that we ought to 'reject' indeterminate propositions (Caie (2012) p.3 n.8). But Kripke seems fairly explicit in the passages just cited that his view is to be interpreted as one on which there are no paradoxical or indeterminate propositions that are candidates for such rejection.

I want to flag these positions for the sake of completeness. I am skeptical of them, for reasons familiar from the literature on the semantic paradoxes. And I am wary of an extension of such views to the current case (an extension which needn't be endorsed by any of the authors just cited). But there is no space to discuss these issues here in adequate detail.[9] In any event, we don't need to lean on the relevant worries here. There are separate problems with Caie's streamlined argument that run orthogonal to the issue of expression failure. Let's grant that there is a proposition expressed by (2). Then, I contend, Caie's argument is valid only if that proposition has a structure that we have reason to think no proposition has.

To see this, let's begin by focusing on the more standard theories, like Fregean or Russellian views, according to which propositions have something akin to linguistic structure. Caie's argument requires that if Ada believes that she doesn't believe the proposition expressed by (2), she *ipso facto* believes the proposition it expresses. What would a structured proposition have to look like for this to be true? The answer is that it would need something like infinite nested structure.

Focus on the Russellian who thinks that propositions are built from objects, properties, and relations (things will be analogous for the Fregean). What should the proposition expressed by (2) look like for such a theorist?

(2) Ada doesn't believe the proposition expressed by (2).

Something roughly consisting of the following four elements: an entity or relation corresponding to negation, a binary belief relation, Ada, and a structured logical entity corresponding to the definite description. This latter construct would itself contain the expression relation, the property of being a proposition, and the English sentence type "Ada doesn't ...".

Let's have "$(\iota x)(\ldots x \ldots)$" designate whatever finitely structured entity corresponds to the use of the definite description. Then the proposition I've just described looks something like the following:

[9] The basic problem for claiming that (2) fails to express a proposition is that someone ignorant of the semantics of (2) could seemingly come to believe that 'what it says' is true by accident. The relevant true belief reports seem to require a propositional object. See Burge (1984) pp.10–2, Horwich (1998) p.43, Soames (1999) pp.193-4, Field (2008) pp.132–3, and Schroeder (2010) pp.284–5, the first of whom applies the argument to epistemic paradox itself. I sympathize with the argument, but I think the case is more complicated than these authors seem to recognize. See Glanzberg (2003) §6, [citation omitted] for discussion of some worries and complexities.

⟨Ada doesn't believe (ɿx)(...x...)⟩

Call this structured proposition p_1.

Could p_1 be the proposition expressed by (2) according to Caie? No. The problem is that if Ada believes that she doesn't believe p_1, she would not *ipso facto* believe p_1. The proposition *that Ada doesn't believe* p_1 would be a structured entity consisting of *two* instances of the believing relation. Believing that proposition involves doing something different from believing a proposition structured using a single belief relation. The two beliefs just considered have two different structured propositional objects.

The commitment that p_1 isn't expressed by (2) should be confusing to the Russellian. After all, the structured proposition p_1 (or something relevantly similar) exists on the Russellian picture, since all of its constituents do. Why would it not be what (2) expresses? After all, its propositional structure perfectly parallels (2)'s linguistic structure.

I'm not sure how one would answer this question. But I do know what kind of proposition (2) should instead express for the Russellian were it to behave as needed for Caie's argument. It would have to be such that the proposition that one doesn't believe it is identical with that very proposition. The trick to construct such a proposition is to exploit infinities (or at least 'self-containment'). In particular, the proposition should be structured in something like the following way.

⟨Ada doesn't believe ⟨Ada doesn't believe ⟨Ada doesn't believe ⟨. . .⟩⟩⟩

The ellipsis is meant to indicate an infinite descent of nested attributions of non-belief. Call this hypothesized proposition p_2. Note that if one embeds p_2 within another proposition to the effect that Ada doesn't believe it, you will just get p_2 back. The proposition *that Ada doesn't believe* p_2 is identical with p_2 itself. Hence believing the former is believing the latter. Or at least it would be if p_2 existed.

If propositions have broadly linguistic structure, then Caie's streamlined argument presupposes both that a proposition like p_2 exists and that it is expressed by (2). But neither claim seems true.

Let me begin with the second issue, of whether (2) expresses a proposition like p_2, starting with some helpful remarks of Russell. Russell was troubled by questions about infinitely structured propositions while developing the theory of denoting concepts that gradually matured into his celebrated 'new' theory of denoting in Russell (1905). Russell's decided view was sober: "...I see no possible way of deciding whether propositions of infinite complexity are possible or not; but this at least is clear, that all the propositions known to us (and, it would seem, all propositions that we can know) are of finite complexity."[10] Russell's thought was that although we are able to think about infinite classes, for example, this is not done by grasping all members of the class at once, as we might for singular propositions about objects of acquaintance. Instead we think about the members of that class indirectly by characterizing their commonalities, often with the help of logical tools. Early on, Russell's tools for that purpose were denoting concepts, but later they were essentially replaced with the familiar logical mechanisms of quantification.

Russell's view here seems right. It is unclear what reasons we have to posit infinitely structured propositions, or to ban their existence. But (and this is the relevant point for our dialectic) we do have some very strong reasons to think they are not expressed by simple sentences like (2). We never need to posit infinitary propositions as direct objects of cognition, since indirect characterizations using quantifiers or other logical tools seem to suffice. And we have an obvious reason against positing infintary objects of our cognition: it seemingly requires us to cognitively grasp infinitely many things at once, when we have only finite cognitive resources. If (2) expresses a proposition at all, it expresses one that a finite mind can cognize. (For example, Rohan might think that (2) expresses the proposition that snakes have wings, and believe that Ada doesn't believe that proposition. Accordingly, Rohan believes that Ada doesn't believe the proposition expressed by (2)—that is, *he* believes the proposition that (actually) is expressed by (2).[11] Rohan doesn't need extraordinary cognitive faculties to do this.)

So, (2) doesn't express an infinitely structured proposition like p_2. But this leaves a window for Caie's argument to resurface. After all, in discussing Russell I sided with him in conceding that infinite propositions might exist, even if not expressed by (2). For example, infinitely structured propositions could

[10] Russell (1903) §141.

[11] One could deny that Rohan, in his state of ignorance, grasps the same proposition one would when one knows (2)'s semantics. But then one would lose what is probably the best argument that (2) actually expresses a proposition—namely that it can be 'accidentally' believed owing to such ignorance. See the discussion and citations in 9.

well make sense in the abstract, roughly as, or as analogous to, certain infinitary set-theoretic constructions. As long as there are some such propositions, won't the argument get going again?

Likely not. Even if we posit (say) propositional infinitary conjunctions, it would still be a noteworthy conceptual leap from their existence to the infinitely descending *nestings* of propositions like p_2. Such infinite nestings do not even correspond to any well-founded set theoretic construction. Perhaps an infinitary propositional conjunction might be graspable by an 'infinite mind', or subsist independently of any minds. But infinite nested propositions like p_2 are bizarre enough that I think we should be very suspicious of their existence without strong arguments that positing them is required for propositional theorizing. Even largely unfettered procedures for constructing infinitary propositions won't generate p_2. This gives us reason to reject their existence, even when we are freed from concerns about the cognitive limitations of finite minds.[12]

As noted, analogous worries arise for the Fregean. Will the argument fare any better if propositions are unstructured sets of possible worlds? There are two reasons this retreat is unhelpful. First, there is again going to be a pressing substantive worry about grounds for positing the requisite proposition. Is there a set of worlds W, such that Ada counts as not believing the proposition true at worlds W if and only if she is in some world in W? I find it difficult to see what speaks in favor of such a set of worlds existing. Perhaps not much can be said against it, though, beyond that it is a little bewildering owing to the circularity.

But there is a second, much deeper dialectical worry. The key problem with the possible worlds framework is the problem of logical omniscience: modeling propositions with possible worlds has all agents (including mundane

[12] I've been critical of the view that there are propositions which contain themselves as parts. A noteworthy opposing line of thought is given by Barwise & Etchemendy (1987), which engages with the liar paradox by treating liar-like propositions with the resources of non-well-founded set theory. Their discussion does, I think, help reveal that the view that propositions may contain themselves as parts is not simply *incoherent*. But I do not find in their work positive motivations for accepting the existence of the relevant propositions. Liar-like propositions that are troubling enough, and seemingly sufficient to model the contents of ordinary speech and thought, can be found without using the resources of non-well-founded set-theory, and instead by exploiting the ability of propositions to talk about themselves 'indirectly' through the use of descriptive material, or quantificational resources, in the ways I've been discussing above.

beings such as ourselves) believing all entailments of all their beliefs, by fiat.[13] I don't think this problem is necessarily insuperable for the view. But something has to be said about it. And what is said is particularly relevant to our puzzle since we are debating precisely whether *Evidence* is the source of our problems. Adopting the possible worlds framework in this setting without speaking to the problem of logical omniscience is clearly just to beg the question, since it could generate paradox 'without *Evidence*' only by assuming that all agents satisfy *Evidence*'s prescriptive demands as a conceptual necessity. There is a very serious worry that what is said about the problem of logical omniscience will bring us closer to something like the Russellian view, reinstating my earlier worries. At any rate, it is clear, I think, that the burden of proof would be on opponents to show otherwise.

I've maintained that Caie's argument tacitly presupposes the existence of propositions with something like infinite nested structure. But if this is so, where was this assumption made in the argument from (a)–(f)? Which premise or inference should we reject? Probing this question is illuminating. The fault turns out to lie in the inference from (b) to (c).

(b) Ada doesn't believe the proposition expressed by (2).

(c) Ada believes the proposition expressed by "Ada doesn't believe the proposition expressed by (2)".

In my argument I claimed this was secured by *Transparency*, following Caie who labels a biconditional linking (b) and (c) a transparency assumption.[14] But this was an extremely subtle logical sleight of hand. The general principle invoked seems to be something like the following, letting "ρ" be shorthand for "the proposition expressed by":

(T) if A satisfies *Transparency*, and S names a sentence:

$$A \text{ doesn't believe } \rho S \leftrightarrow A \text{ believes } \rho\text{“}A \text{ doesn't believe } \rho S\text{”}.$$

But as a general principle, (T) is false.

To see why, consider sentence (3).

[13] See, e.g., Stalnaker (1984) for a helpful discussion of the problem.

[14] Caie (2012) p.12, line (7). I've replaced Caie's "α" with my "Ada" and his "(*)" with my "(2)".

(3) 2+2=5

Suppose Tal, who satisfies *Transparency*, is peering over my shoulder as I write. Tal knows basic arithmetic, and accordingly has no false belief that 2 and 2 make 5. So:

(I) Tal doesn't believe the proposition expressed by (3).

But Tal's eyesight is bad. He reads (3) well enough, but sees the final numeral "5" as a "4". He accordingly believes that (3) expresses the truth that 2 and 2 make 4. And he believes that he believes that (3) expresses a truth. In other words:

(II) Tal believes the proposition expressed by "Tal believes the proposition expressed by (3)."

If you asserted the sentence mentioned in (II), Tal would understand it and assent to it, for example. But by (I) and (T), we have:

(III) Tal believes the proposition expressed by "Tal doesn't believe the proposition expressed by (3)"

So Tal is driven into inconsistency. But surely this is wrong! There's no reason to think that Tal's beliefs must be inconsistent—he just has a little trouble seeing. What has happened?

The problem is that the definite description in "*A* believes the proposition expressed by *S*" only has what is effectively a *de re* reading.[15] For example, even with as favorable a context as possible, one can't hear the second sentence in (4) as a truth about our story.

(4) Tal thinks (3) expresses the proposition that 2 and 2 make 4, which Tal believes. So, Tal believes the proposition expressed by (3).

The last sentence can only claim, falsely, that Tal believes whatever proposition (3) actually expresses (namely, that 2 and 2 make 5) not, truly, that Tal believes the proposition he takes (3) to express (that 2 and 2 make 4).

But when the definite description becomes part of a subordinate clause in "*A* believes that *A* believes the proposition expressed by *S*", it acquires both a

[15] "Effectively" because the *de re/de dicto* distinction may not even apply to this sentence. Note that "believes" in this sentence doesn't, e.g., take a clausal complement.

de re and *de dicto* reading (with respect to the outermost "believes"). There is, for example, a quite natural true reading of the final sentence in (5) in context (though there is a false *de re* reading as well).

(5) Tal thinks (3) expresses the proposition that 2 and 2 make 4. And Tal believes that he believes that proposition. So, Tal believes that he believes the proposition expressed by (3).

The natural, true reading is roughly: Tal believes that the proposition he takes to be expressed by (3)—that 2 and 2 make 4—is among those he believes.

The problem with principle (T) is then the following. The description ρS on the left-hand side of the biconditional must effectively be read *de re*—as I said, there's no other option. But the quoted expression on the right-hand side 'traps' ρS, effectively forcing a *de dicto* reading.[16] But the left-hand side of the biconditional, coupled with the assumption of transparency, won't ensure that *de dicto* claim is true. Instead it will ensure the iterated *de re* belief ascription is true. In our case, for example, an instance of the biconditional is simply false: (I) is true but (III) is false.

A related issue arises in Caie's deduction. Seeing things from the Russellian's perspective may help, though it's worth stressing that the ensuing exercise is a heuristic: the general point about (T) does not turn on the truth of the Russellian, or more generally structured, view of propositional content.

We start by assuming (2)/(b).

(b) Ada doesn't believe the proposition expressed by (2).

Note, (b) expresses the following Russellian proposition.

$$\langle \text{Ada doesn't believe } (\iota x)(\ldots x \ldots) \rangle$$

Now, this proposition doesn't give us enough information to fully understand what Ada's transparent beliefs allow her to know. It says that Ada doesn't believe something, and it *describes* the proposition she doesn't believe. It's not as if we're claiming Ada has two beliefs: a belief in the proposition described and a belief with the description as object! The latter has no sense (which is why the description must effectively be read *de re*). So to really know what Ada knows

[16] "Effectively", since it is only forcing an equivalent of the *de dicto* reading, by exploiting quotation.

by reflection, we should settle what proposition satisfies the description. The proposition in question is the very proposition expressed by (b). So we have:

Ada doesn't believe: ⟨Ada doesn't believe (ιx)(...x...)⟩

Now that we've specified the proposition she doesn't believe, we can figure out what Ada knows by reflection on her own mental state. If Ada satisfies *Transparency*, what follows is:

Ada believes: ⟨Ada doesn't believe ⟨Ada doesn't believe (ιx)(...x...)⟩⟩

But this is not what Caie infers. He infers (c).

(c) Ada believes the proposition expressed by "Ada doesn't believe the proposition expressed by (2)".

That is:

Ada believes: ⟨Ada doesn't believe (ιx)(...x...)⟩

This is simply not the proposition that Ada believes by an application of *Transparency*. Note, however, that it is *close* to following from *Transparency*. The proposition that Ada actually believes by *Transparency*, above, along with information about the denotation of the description, *entails* the proposition Caie has Ada believe.

So Caie's transparency assumption is actually smuggling in just a little more than genuine transparency. It would accordingly be best to split Caie's single transparency biconditional into two biconditionals: one for genuine transparency and one for the 'remainder'. A genuine transparency principle, if it involves descriptions of propositions at all, should bridge first and second order *de re* belief.

$\rho(2)$ is such that Ada doesn't believe it $\leftrightarrow$
$\rho(2)$ is such that Ada believes that Ada doesn't believe it

If there is some proposition, which happens to meet the description "$\rho(2)$", that Ada doesn't believe, then *that* proposition is such that she'll believe that she doesn't believe it. Note that either instance of "$\rho(2)$" can be freely substituted *salva veritate* with any term that refers to the same proposition.

To get Caie's biconditional, we must add a principle bridging second order *de re* belief, and second order *de dicto* belief.

$\rho(2)$ is such that Ada believes that Ada doesn't believe it $\leftrightarrow$
Ada believes ρ"Ada doesn't believe $\rho(2)$"

This in part claims: if a proposition p which happens to fall under the description "ρ(2)" is such that Ada believes that she doesn't believe p, then Ada believes that she doesn't believe whatever proposition falls under the description $\rho(2)$. Put another way:

$\rho(2)$ is such that Ada believes that Ada doesn't believe it $\leftrightarrow$
Ada believes that Ada doesn't believe $\rho(2)$

On the left-hand side "$\rho(2)$" is substitutable *salva veritate* with any term referring to the same proposition. On the right-hand side it cannot. It is no longer merely describing the propositional object of Ada belief—it's now a description under which Ada is thinking about a proposition.

This second principle is in no way a transparency principle. Assuming it without defense is obviously to beg the question. So what could ground it?

As already noted, the biconditional could be grounded in an inference Ada has performed. We've already seen this in the recent discussion of the Russellian view. If the left-hand side of this biconditional is true, Ada believes that she doesn't believe some proposition p. But she's also in a position, on reflection, to see that $p = \rho(2)$. And it is short step from there for her to infer that she doesn't believe $\rho(2)$, so that the right-hand side becomes true.

But to suppose Ada is *forced* into irrationality because the second biconditional is true in this way is to suppose that Ada is *forced* by the dictates of rationality to perform the relevant inferences. But to suppose that would directly undermine the point of the new argument, which was to do without the assumption of *Evidence*. These are the very inferences that I earlier argued are not rationally required for Ada to perform. Caie claimed his argument was going to bypass the reliance on rationally requiring these inferences, and so avoid the controversy raised by *Evidence*. But if we pursue this route, *Evidence* has merely been surreptitiously incorporated into a principle labelled as a transparency assumption.

Given this, I see only one option left: to take the truth of the biconditional to be secured not by inferential moves, but by the metaphysics of the proposition expressed by (2). That is, we should see this principle as constituting the assumption that the proposition expressed by (2) is such that, if Ada believes that she disbelieves it, she *ipso facto* believes it, owing only to the nature of the proposition in question. This, too, could ensure the truth of the second biconditional. But, as I've said, there is very strong evidence that (2) expresses no proposition of the required sort. And, more strongly, there are grounds to think that no proposition with these odd features exists, independently of the question of whether (2), or any other natural language sentence, expresses it.

So we are back to the drawing board. *Evidence* is an ineliminable assumption used in formulating epistemic paradox. I've already argued we have plenty of grounds to safely reject *Evidence* in resolving the tensions the paradox creates (perhaps the strongest grounds coming from epistemic paradox itself). So that paradox does not drive us to the position Caie promotes.

2 Epistemic Paradox and Epistemic Limits

On the view I'm proposing, what does the epistemic paradox teach us? I've explained why I think it doesn't reveal that we are rationally required to get into indeterminate doxastic states. Does it teach us we should reject *Evidence*? In a sense. As discussed earlier, I don't think this is something that epistemic paradox uniquely teaches. The ways we are rationally required to expand our beliefs are complex, depending on foreseeable utility, metaphysical constraints on inferential transitions and, of course, our epistemic limitations. I think a principle describing these rational requirements at a suitable level of generality would have been too weak to generate epistemic paradox to begin with.

Epistemic paradox, on my view, teaches us relatively little about rational belief. But it does, I think, teach us something rather striking about *knowledge*.[17] Indeed, it may prove to be the missing half of a general result on the structural limits of knowledge.

The first half of those structural limits are provided by Fitch's paradox.[18] This puzzle begins by noting that (6) could be true of Ada.

(6) Ada snored last night and it is never known that Ada snored last night.

But if (6) is true, we can show by *reductio* that (6) expresses an unknowable truth—never known by anyone, at any time, in any world. If (6) were known, both conjuncts would be. But if someone were to know the first conjunct, the

[17] It may well teach us something about rational *credences*: see Caie (2013). I won't be able to discuss these issues here.

[18] Fitch (1963).

second conjunct would clearly be false. But then the whole conjunction in (6) would be untrue, and hence not the sort of thing that could be known.

Whether this *reductio* goes through is contested. But even if it does, there are two things that Fitch's paradox clearly does not show (even for beings at worlds with unknown truths): that omniscience is impossible or that there is a true proposition p such that it is impossible to know *whether* p is true. The deduction can't show the former in part because it requires a premise about unknowns. And it can't show the latter because the deduction is silent, for any unknowable p, on whether it is possible to know p's negation.[19]

Epistemic paradox, however, does show something very close to both claims. Before saying why, first briefly note that although Caie's streamlined paradox could conceivably be resisted by denying that (2) expresses a proposition, this move is significantly harder to maintain for other versions of the paradox that undisputedly require appeal to *Evidence*. Consider (7).

(7) At t, Ada does not believe that (7) expresses a true proposition.

It is very challenging to maintain that (7) doesn't express a proposition in large part because one can no longer lean on the semantics of definites to explain the presence of semantic defect. For these reasons (7) is also unhelpful in bypassing *Evidence* in generating paradox. But we've already given up that goal.

Now, if this is right, (7) expresses a proposition that Ada is rationally required to be agnostic about at t insofar as she satisfies *Transparency* and understands its semantics. If she believes or disbelieves the proposition expressed by (7) under the relevant conditions, her actual beliefs entail a contradiction—namely (7) and its negation. None of the excuses I ran through in objecting to *Evidence* explain away the irrationality of her state. The entailed contradiction here is foreseeable. And she can't claim the proposition is a trivial irrelevancy she doxastically ignored—she didn't ignore it. If, by contrast, she's agnostic about the proposition expressed by (7), her beliefs don't entail a contradiction. Instead they entail a Moore-paradoxical claim of the form "q and Ada doesn't believe q". But that's just the thing about Moore-paradoxical statements: like contradictions, they're irrational believe or assert (neither of which Ada will have done); but unlike contradictions, there's no reason they can't be true. In this case, Ada's beliefs merely entail the relevant truth. There's no special reason to count Ada irrational here (at least, if we're granted *Possibility*)—she's

[19] See Melia (1991), Williamson (2000) §12.4.

done the best she epistemically can.

For related reasons (7) presents a limited block to Ada's omniscience: it prevents Ada from being omniscient at *t*. This consequence is shared on Caie's view. Let's review why. Grant that (7) expresses a proposition *p* that is true if and only if Ada doesn't believe (7) expresses a truth at *t*. If Ada merely believes at *t* that (7) expresses a truth, it won't, so Ada's belief can't constitute knowledge of whether *p*. If Ada merely believes (7) doesn't express a truth, it will express a truth, so her disbelief can't constitute knowledge of whether *p*. If she believes neither *p* nor its negation, she won't know whether *p* for lack of trying. If it's indeterminate whether she believes *p*, she hovers between two or more states, none of which constitute knowledge. In such a case, indeterminate believing doesn't entrain indeterminate knowing, but determinate lack of knowledge. Finally if she believes both *p* and its negation, *p*'s negation will be true, but it is dubious that her belief in *p*'s negation could constitute knowledge. For example, Ada's opinions about *p* are not at all reliable, they are irrational, and so on.[20] Note this argument doesn't require *Evidence*, *Possibility*, or *Transparency*. It only requires that (7) express a proposition that is true just in case Ada doesn't believe at *t* that (7) expresses a truth, and *Consistency*.[21]

Note that for this to be a claim about metaphysical possibility we have to assume, as I will, that the language containing (7) is abstract, and that (7) and its actual semantics can be thought about even at worlds where the language of which it is a part is never used.[22] But once we grant this, the point can generalize. As long as for each point in time t' there is an abstract expansion of English that contains a way of referring to that point in time, there will be an abstract sentence expressing a proposition (in part about that abstract language) whose truth or falsity is unknowable for Ada at t'. And if, for each thinker, we can find an expansion of English with a name for her, then there is a corresponding set of such temporally indexed sentences for each cognizing being. Thus the failure of omniscience isn't temporary, but eternal. It isn't personal, but

---

[20] If you doubt this last claim then at least *rational* omniscience is impossible, or 'omniscience' defined so that all of one's beliefs are known. See Grim (1983) §1 for a discussion.

[21] The basic idea here has been put forward in Grim (1983) p.267–8, though Grim's argument (reasonably) neglects worries about indeterminate belief, and would only show the impossibility of eternal omniscience. Though I've conceded that (7) expresses a truth-evaluable proposition, I'm not actually sure that should actually be conceded either. Whether it does may depend on what parallels we find between (7) and the paradox of the knower (see Kaplan & Montague (1960)), which replaces "believes" with "knows".

[22] See, e.g., Lewis (1975).

universal. If it is the essence of God to be omniscient, not only does God not actually exist, but he couldn't possibly exist. Fitch's deduction, even if it is accepted, cannot show anything like this.

Fitch's paradox purports to show that if there is a true proposition unknown by anyone at any time, then there is a single true proposition that is unknowable by anyone at any time:

$$(\exists p)(p \wedge (\forall A)(\forall t)(\Box\neg K^t_A(p))$$

Epistemic paradox, by contrast, seems to show (with no assumptions about what actually is known) that, for any agent, time, and world, we can find a proposition such that the agent doesn't know whether that proposition holds at that time:

$$(\forall A)(\forall t)(\Box(\exists p)(\neg K^t_A(p) \wedge \neg K^t_A(\neg p))$$

This is the sense in which epistemic paradox may ground one half of a complementary set of structural restrictions on knowledge.

Though the structural limit on knowledge that precludes omniscience is certainly of theoretical interest, surely it is not of much practical interest. Most of us never dreamed of such vast knowledge. For finite beings like us, as I noted in discussing *Evidence*, very many propositions—many of them truths—really aren't worth the trouble of thinking about. I suggest that, for us at any rate, the propositions figuring in our epistemic paradoxes should be among them.

References

Barwise, Jon & John Etchemendy. 1987. *The Liar: An Essay on Truth and Circularity*. Oxford University Press. [12]

Boghossian, P. 2014. "What is Inference?" *Philosophical Studies*, vol. 169 (1): 1–18. [4]

Broome, John. 2014. "Comments on Boghossian." *Philosophical Studies*, vol. 169 (1): 19-–25. [4]

Burge, Tyler. 1978. "Buridan and Epistemic Paradox." *Philosophical Studies*, vol. 34 (1): 21–35. [2]

—. 1984. "Epistemic Paradox." *Journal of Philosophy*, vol. 81 (1): 5–29. [2], [9]

Byrne, Alex. 2005. "Introspection." *Philosophical Topics*, vol. 33 (1): 79–104. [5]

—. 2011. "Knowing That I Am Thinking." In *Self-Knowledge*, Anthony Hatzimoysis, editor, 105–124. Oxford University Press. [5]

Caie, Michael. 2012. "Belief and Indeterminacy." *Philosophical Review*, vol. 121 (1): 1–54. [1], [2], [7], [8], [13]

—. 2013. "Rational Probabilistic Incoherence." *Philosophical Review*, vol. 122 (4): 527–575. [18]

Conee, Earl. 1987. "Evident, but Rationally Unacceptable." *Australasian Journal of Philosophy*, vol. 65 (3): 316–26. [4], [6]

Field, Hartry. 2008. *Saving Truth From Paradox*. Oxford University Press. [9]

Fitch, F.B. 1963. "A Logical Analysis of Some Value Concepts." *The Journal of Symbolic Logic*, vol. 28 (2): 135–142. [18]

Gibbard, Allan & William Harper. 1978. "Counterfactuals and Two Kinds of Expected Utility." In *Foundations and Applications of Decision Theory*, 125–162. D. Reidel. [3]

Glanzberg, Michael. 2001. "The Liar in Context." *Philosophical Studies*, vol. 103 (3): 217–251. [8]

—. 2003. "Against Truth Value Gaps." In *Liars and Heaps: New Essays on Paradox*, J.C. Beall, editor, 151–194. Clarendon Press, Oxford. [9]

Grim, Patrick. 1983. "Some Neglected Problems of Omniscience." *American Philosophical Quarterly*, vol. 20 (3): 265–77. [20]

Horwich, Paul. 1998. *Truth*. Clarendon Press, Oxford. [9]

Kaplan, David & Richard Montague. 1960. "A Paradox Regained." *Notre Dame Journal of Formal Logic*, vol. 1: 79–90. [20]

Kripke, Saul. 1975. "Outline of a Theory of Truth." *The Journal of Philosophy*, vol. 72 (19): 690–716. [8]

Lewis, David. 1975. "Languages and Language." In *Minnesota Studies in the Philosophy of Science*, vol. 7, 3–35. University of Minnesota Press. [20]

Melia, Joseph. 1991. "Anti-Realism Untouched." *Mind*, vol. 100 (3): 341–342. [19]

Neta, Ram. 2013. "What is an Inference?" *Philosophical Issues*, vol. 23 (1): 388–407. [4]

Priest, Graham. 1985. "Contradiction, Belief and Rationality." *Proceedings of the Aristotelian Society*, vol. 86: 99–116. [3]

—. 2002. "Rational Dilemmas." *Analysis*, vol. 62 (1): 11–16. [3]

—. 2006. *In Contradiction: A Study of the Transconsistent*. Oxford University Press. [3]

Richter, Reed. 1990. "Ideal Rationality and Hand Waving." *Australasian Journal of Philosophy*, vol. 68 (2): 147–56. [4], [5], [6]

Ross, Jacob. 2010. "Sleeping Beauty, Countable Additivity, and Rational Dilemmas." *Philosophical Review*, vol. 119 (4): 411–447. [3]

Russell, Bertrand. 1903. *Principles of Mathematics*. Routledge. [11]

—. 1905. "On Denoting." *Mind*, vol. 14 (56): 479–493. [11]

Schroeder, Mark. 2010. "How to Be an Expressivist About Truth." In *New Waves in Truth*, 282–298. Palgrave Macmillan. [9]

Soames, Scott. 1999. *Understanding Truth*. Oxford University Press, Oxford. [9]

Sorensen, Roy. 1987. "Anti-Expertise, Instability, and Rational Choice." *Australasian Journal of Philosophy*, vol. 65: 301–15. [4]

Stalnaker, Robert. 1984. *Inquiry*. MIT Press. [13]

White, A. R. 1971. "Inference." *Philosophical Quarterly*, vol. 21: 289–302. [4]

Williamson, T. 2000. *Knowledge and Its Limits*. Oxford University Press. [4], [19]

Wright, Crispin. 2014. "Comments on Paul Boghossian, "What is Inference"." *Philosophical Studies*, vol. 169 (1): 27–37. [4]