IT IS NEVER RATIONAL FOR ANYONE TO BELIEVE THEY DON'T KNOW THE LOGICAL TRUTH

Luis Rosa
Washington University in St. Louis

Forthcoming in *Thought: A Journal of Philosophy*
Preprint—please cite the published version

Abstract: Let T be any logical truth. Does the subject know that T? It is not rational for any subject to believe that *they don't*, whoever they are. Similarly, it is not rational for them to believe that *their evidence doesn't support T*, and it is not even rational for them to believe that *they don't believe that T*. It is not rational for anyone anywhere at any time to believe that *they don't know that T*. Such are the conclusions arrived at in this paper.

1

It seems clear that there are logical truths we don't know. And it seems furthermore rational for us to believe that there are such truths. A few plausible assumptions suggest that these appearances are misleading, however.

Let T be any logical truth, and let S be any arbitrary subject. In order to get to the conclusion that it is not rational for S to believe that *they don't know that T*, we use the following premises:

(1) For any p, p is logically equivalent to $(T \wedge p)$.

(2) For any S, p, if it is rational for S to believe that p then, for any q that is logically equivalent to p, it is possible for S to rationally believe that q.

(3) For any S, p, it is impossible for S to rationally believe that ($p \wedge$ *S doesn't know that p*).

Premise (1) should be familiar to those who have been exposed to standard (classical) logic. Take any arbitrary proposition p. Suppose p is true. Since T is true, it follows from our supposition that $(T \wedge p)$ is true. Now suppose that $(T \wedge p)$ is true. It follows that p is true. Therefore $(T \wedge p)$ follows from p and vice-versa: they are logically equivalent.

The concept of rationality in the antecedent of the conditional in (2) is a concept of *ex ante* rationality: it can be rational for S to believe that p even though S doesn't actually believe that p.[1] The concept of rationality in its consequent, however, is a concept of *ex post* rationality. It says, then, that it is at least possible for S to *rationally believe* that q, given that q is logically equivalent to p and it is rational for S to believe that p.

We can think of the relevant kind of possibility as 'metaphysical' or genuine possibility. A possibility in this sense is not just something that *we take* to be possible or that we conceive of as being possible—but is something that is indeed possible. The concept of rationality in (3) is also a concept of ex post rationality, and the kind of possibility at stake there is the same as the one from the consequent of (2).

Any notion of rationality that makes (2) and (3) true—say, rationality as proper responsiveness to the evidence—will generate the consequences explored below. And maybe there are notions of rationality that do not make (2) and (3) true. But how relevant or useful such notions of rationality are, and how they compare to notions of rationality that make (2) and (3) true, is not something we can hope to decide here.

It is important to emphasize that in (3) the content ($p \wedge$ *S doesn't know that p*) must be presented to S in such a way that S recognizes the second conjunct as referring to themselves.[2] Furthermore, all the time indexes supplied by our context of utterance should be the same throughout all the time-slots in (3): it is impossible for S to rationally

[1] If we were to phrase our premises in terms of *epistemic justification*, then the antecedent in (2) would be concerned with what is called *propositional* justification, its consequent with *doxastic* justification. (3) is also concerned with doxastic justification.

[2] It is not necessarily irrational for you, for example, to believe that ($p \wedge$ *S doesn't know that p*) when you don't recognize S as being yourself. These are supposed to be cases of 'Moore-paradoxical' propositions—see Moore (1993), de Almeida (2001) and Williams (2015). Notice however that there are philosophers who dispute the claim that Moorean assertions are always defective or absurd—see for example Turri (2010) and Pruss (2012).

to believe *at t* that ($p \wedge$ *S doesn't know that p at t*).[3] Time indexes are left implicit in the interest of avoiding clutter, same as in (2). We can assume uniformity of time indexes for all other sentences used in the demonstration to follow—whenever there is a time-slot, just plug in the same time index in it. (We quantify over any particular time t and run the proof under that quantification, for any time t).

We can now make a *reductio* of the assumption that it is rational for you to believe that *you don't know that T*:

(4) It is rational for you to believe that *you don't know that T* (Supposition for *reductio*).

(5) *You don't know that T* is logically equivalent to ($T \wedge$ *you don't know that T*) (From 1).

(6) It is possible for you to rationally believe that ($T \wedge$ *you don't know that T*) (From 2, 4, 5).

(7) It is impossible for you to rationally believe that ($T \wedge$ *you don't know that T*) (From 3).

(8) It is not rational for you to believe that *you don't know that T* (6, 7, *reductio* of 4).

Suppose someone asks you: do you know whether T? You consider T and cannot make up your mind as to whether T is true or false. So you reply that you don't know whether T. Suppose you unpack that and come to believe that *you don't know that T* (*and you don't know that not-T* either). Reasonable as that may sound, (8) tells us that it is not rational for you to believe that *you don't know that T.*

We can run the same reasoning with respect to any random subject, and conclude on that basis that it is never rational for anyone to believe that *they don't know that T.* If that conclusion is false, however, then one of the premises (1)–(3) must be false. But which one(s)? They all seem plausible—it's a puzzling situation.

[3] It is not necessarily irrational for S to believe at t, for example, that ($p \wedge$ *S doesn't know that p at* t^*) where $t^* < t$.

2

One more or less immediate reaction to the puzzle is that of rejecting (2). But it is hard to find independent grounds to reject it (independent from the puzzle presented above itself).

After all, (2) requires the mere *possibility* that the subject rationally believes a proposition that is logically equivalent to a proposition that is rational for them to believe. Consider some examples. If it is rational for you to believe that *All ravens are black* then, presumably, it must at least be *possible* for you to rationally believe that *All non-black things are non-ravens*. And, if it is rational for you to believe that *Ruth Marcus was a logician*, then it must at least be possible for you to rationally believe that *It is not the case that Ruth Marcus was not a logician*. As a special case, it is rational for you to believe that *you have hands* only if it is possible for you to rationally believe that *you have hands* (every proposition is logically equivalent to itself).

Notice that one cannot raise those objections that are typically raised against so-called epistemic 'closure' principles against (2): that the subject might fail to grasp the connection between the premises and the conclusion, or the subject might lack the cognitive skills to deduce the latter from the former, etc.[4] For what one would need to deny (2) is the *impossibility* that the subject realizes that the premises and the conclusion are so related, or the *impossibility* that the subject has the relevant reasoning skills. In contrast, denying that the subject *actually* fails to realize that the premises and the conclusion are so related, or that the subject *actually* has the relevant skills, can be enough to make trouble for the typical closure principles (for example, the principle that if it is rational for S to believe that p, and p is logically equivalent to q, then it is rational for S to believe that q).

Suppose that it is rational for you to believe some proposition p. p is logically equivalent to $\neg\neg p$, for example (double-negation). Even if *in the actual world* you fail to see the equivalence between p and $\neg\neg p$, or you lack the ability to deduce the latter from the former, there should still be a possible world where you rationally believe that $\neg\neg p$. In that possible world, say, you have the evidence/knowledge that you have in the actual world (where it is rational for you to believe that p) plus whatever it is

[4]See Luper (2020) for an overview of this literature.

that you need to be sensitive to the equivalence relation between those two propositions, or to deduce one from the other.

Why should such a scenario be *impossible*, given that it is actually rational for you to believe that p? (And similarly for other propositions that are logically equivalent to p).

3

Another alternative is to reject (3). What of this strategy?

Here is one way of rejecting this premise. First, the *ex ante* rationality of belief in p for a subject S is a matter of whether S's evidence supports p. If it does, then it is rational for S to believe that p, otherwise it isn't. Second, S can perfectly well be in possession of a total body of evidence that supports not only the two conjuncts p and *S doesn't know that p*, but also their conjunction ($p \wedge$ *S doesn't know that p*).[5] So why couldn't S rationally believe that proposition on the basis of that evidence?

A rival view would say, however, that coherence is also a requirement of rationality, at least *ex post* rationality, if not *ex ante* rationality. And, since belief in the target conjunction is necessarily incoherent, as held by S, (3) remains true.[6]

We won't try to settle this dispute here. Rather, we note that those who reject (3) will similarly have to reject many other 'anti-akrasia' principles such as the following:

(9) For any S, p, it is impossible for S to rationally believe that ($p \wedge$ *S doesn't believe that p*).

(10) For any S, p, it is impossible for S to rationally believe that ($p \wedge$ *S's evidence doesn't support that p*).

(11) For any S, p, it is impossible for S to rationally believe that ($p \wedge$ *it is irrational for S to believe that p*).

---

[5]For some relevant discussion, see Williamson (2014), Worsnip (2018), Lasonen-Aarnio (2020).

[6]This view does not entail the denial of evidentialism about rational belief. For it is possible to think of coherence as a necessary condition for evidential support, or rather as an independent dimension of rationality. See Smithies (forthcoming) for a view of the former sort, Wornip (2018, 2021) for a view of the latter sort.

For, using any of these, together with (1) and (2), we can reach similar conclusions as the one we have reached above involving the logical truth T, namely conclusion (8). For example, (9) will lead to the conclusion that it is never rational for anyone to believe that *they don't believe that T*, (10) will lead to the conclusion that *it is never rational for anyone to believe that their evidence doesn't support T*, etc., through the same kind of *reductio* argument that led us to conclude that it is never rational for anyone to believe that *they don't know that T*.[7]

But it is hard to believe that we should reject all of (3) and (9), (10), (11), etc. Take (11) for example. Rejecting it commits us to saying that S can rationally believe that ($p \wedge$ *it is irrational for S to believe that p*). But then rationality will guarantee the falsehood of the very rational belief! Wasn't epistemic rationality supposed to be truth-conducive? Then how can a belief be rationally held when its rationality prevents its own truth? (If S rationally believes that ($p \wedge$ *it is irrational for S to believe that p*), then it is rational for S to believe that p. But then it is false that it is irrational for S to believe that p.)

The rejection of (9) and (10) and other similar-looking principles will have their own problems, too. Those who reject all of (3) and (9), (10), etc. will have to offer quite a mixed bag of explanations for the failure of each of these principles and deal with a number disparate problems as a result (problems such as the ones presented in the form of questions in the previous paragraph).

This puts considerable pressure on the strategy of attempting to solve our puzzle by rejecting (3). We conclude that this is not an efficient strategy—it solves one instance of the general problem without solving the others.

4

Should we then reject the logical principle in (1)?

[7]Suppose for example that it is rational for you to believe that *you don't believe that T*. The proposition that *you don't believe that T* is logically equivalent to the proposition that ($T \wedge$ *you don't believe that T*). By (2), then, it should be possible for you to rationally believe that ($T \wedge$ *you don't believe that T*). But that is inconsistent with (9). If (9) is true, then, it is not rational for you to believe that *you don't believe that T*.

There would then be some proposition p such that p is not logically equivalent to $(T \wedge p)$, despite the fact that they always have the same truth-value at every logically possible world/model. But then we might reasonably reply that making such a move consists of an attempt to change the concept of logical equivalence, or it confuses logical equivalence for some other relation. (Contrast to *synonymy*: p and $(T \wedge p)$ might be be logically equivalent even though 'p' and '$(T \wedge p)$' are not synonymous sentences, or they do not have the same meaning).

Instead of making such a grandly revisionary move, perhaps we should just accept the conclusions derived from the supposition that T is a logical truth here, counter-intuitive as those conclusions may be. We would then accept, for example, that it is not rational for you to believe that *you don't believe that T*, that *your evidence doesn't support T*, etc., and the same applies to anyone else at any time (with the target propositions being adapted to the varying subjects and times).

These consequences may sound strange at first. Upon closer inspection, however, they are intimately connected to well-known features of the canonical formal systems that are deployed in formal epistemology. In accepting those strange conclusions, it is as if we were taking the testimony of standard epistemic logic and Bayesianism to heart. For standard epistemic logic has it that we know that T at any possible world. That is because T is true at all possible worlds, *ergo* it is true at all possible worlds that are compatible with what the subject knows. Similarly, standard doxastic logic has it that we believe that T at any possible world. Furthermore, both subjective and objective forms of Bayesianism are committed to the idea that the rational credence for us to have in T is the maximal one. For, where Pr is a probability function, $Pr(T) = 1$.

It is in fact intriguing that our premises here made no use of the idealizing assumptions that are constitutive of these formal models of belief, rational credence and knowledge. And yet, again, our premises entail that it is irrational for us to reject certain deliverances of those idealized formal models, even when they are quite literally interpreted (those models 'say' that we know that T, that we believe that T, that it is rational for us to be certain that T). Perhaps that is even stranger than the conclusions that we have arrived at above, and it cries for a deeper investigation that extrapolates from the confines of this paper.

Be that as it may, the jury is now out regarding the puzzling arguments concerning the rationality of higher-order beliefs presented above: should we stick to their negative conclusions or rather reject their premises? The reasons just given favor sticking to those negative conclusions. It is never rational for anyone anywhere to believe that *they don't know that T*, that *their evidence doesn't support T*, etc.

References

de Almeida, Claudio (2001). "What Moore's Paradox is About", *Philosophy and Phenomenological Research* 62(1): 33–58.

Lasonen-Aarnio, Maria (2020). "Enkrasia or evidentialism? Learning to love mismatch", *Philosophical Studies* 177(3): 597–632.

Luper, Steven (2020). "Epistemic Closure", In E. Zalta (ed.) *The Stanford Encyclopedia of Philosophy.*

Moore, G. E. (1993). "Moore's Paradox". In G.E. Moore: *Selected Writings*, ed. by T. Baldwin. London and New York: Routledge.

Pruss, Alexander (2012). "Sincerely Asserting What You Do Not Believe", *Australasian Journal of Philosophy* 90(3): 541–546.

Smithies, Declan (forthcoming). "The Unity of Evidence and Coherence", in Nick Hughes (ed.) *Essays on Epistemic Dilemmas*, Oxford University Press.

Turri, John (2010). "Refutation by elimination", *Analysis* 70(1): 35–39.

Worship, Alex (2018). "The Conflict of Evidence and Coherence", *Philosophy and Phenomenological Research* 96(1): 3–44.

Worsnip, Alex (2021). *Fitting Things Together: Coherence and the Demands of Structural Rationality*, Oxford University Press.

Williams, John N. (2015). "Moore's Paradox in Thought", *Philosophy Compass* 10(1): 24–37.

Williamson, Timothy (2014). "Very Improbable Knowing", *Erkenntnis* 79(5): 971–999.