THE JOURNAL OF PHILOSOPHY
VOLUME CXIX, NO. 11, NOVEMBER 2022

AMBIGUOUS STATEMENTS ABOUT AKRASIA*

Much attention has been given lately to the question of whether it can be rational for one to be akratic. In some versions of the problem, the question is whether there are situations where it is rational for one to believe that *p* while it is at the same time rational for one to believe that *it is not rational for one to believe that p*. Many other negative assessments regarding one's epistemic position with respect to *p* are candidates for the content of the higher-order belief that disavows the lower-order belief.[1]

In this paper, I concentrate on the question of whether it can be rational for a subject to believe that *p* while it is at the same time rational for her to believe that *her evidence does not support p*, regardless of whether the subject goes on and believes those things.[2] All the views that I will discuss here assume that what is rational for a subject to believe depends, perhaps among other things, on what her evidence is and what it gives support to. In particular, it is assumed that it is rational for a subject to believe that *p* only if her evidence supports *p* (her *total* evidence, or all her evidence that bears on the question of whether *p* is true—I will leave these qualifications implicit from now

* This paper was presented at the Brown Bag Seminar at CONCEPT (University of Cologne) and at the First Rhine-Ruhr Epistemology Meeting. I thank the audiences on both occasions for their valuable feedback. I would also like to thank Matthew McGrath for his feedback when I was sketching the first draft of this paper. Finally, I would like to thank the Alexander von Humboldt Foundation for funding this research.

[1] For some of the discussion, see Sophie Horowitz, "Epistemic Akrasia," *Noûs*, XLVIII, 4 (2014): 718–44; Clayton Littlejohn, "Stop Making Sense? On a Puzzle about Rationality," *Philosophy and Phenomenological Research*, XCVI, 2 (2018): 257–72; and Maria Lasonen-Aarnio, "Enkrasia or Evidentialism? Learning to Love Mismatch," *Philosophical Studies*, CLXXVII, 3 (2020): 597–632.

[2] For the idea that higher-order evidence affects *ex post* rationality or doxastic justification without necessarily affecting *ex ante* rationality or propositional justification, see Paul Silva, "How Doxastic Justification Helps Us Solve the Puzzle of Misleading Higher-Order Evidence," *Pacific Philosophical Quarterly*, XCVIII, S1 (2017): 308–28.

0022-362X/22/1911/581–601

on). I will hold onto that assumption throughout the paper, and I will not discuss views according to which it can be rational for a subject to believe that *p* even though her evidence *does not* support *p*: the evidence is neutral about whether *p* or, worse, it gives support to ¬*p*.[3] Clarifications about the notions of evidence and evidential support will be made in due course.

The points that I want to make here have to do with the propositional contents of ascriptions of rationality to hold higher-order beliefs of the relevant sort. They are supposed to explain how some epistemologists can be led to think that there are situations where it is rational for a subject to believe that *p* and at the same time rational for her to believe that *her evidence does not support p*, whereas other epistemologists are led to think that such situations are not possible. The intended explanation is in this way supposed to (a) present us with a rational reconstruction of the target epistemological views, or show how one can arrive at those views through perfectly good (deductive) reasoning, and (b) show that those apparently conflicting epistemological assessments can both be true at the same time.

Unsurprisingly, what makes (b) possible is the fact that 'it is rational for the subject to believe that *p* and it is at the same time rational for her to believe that *her evidence does not support p*' means different things in the mouths of each of the opposing parties. And that is not because those parties mean different things by 'rational' or even by 'supports'. They can be in perfect agreement about *that*. Rather, the relevant difference in meaning lies in the last bit of the quoted sentence, namely, the fragment 'to believe that *her evidence does not support p*'. The italicized part stands for a proposition, or a possible content of a doxastic attitude—but there is more than one way of establishing what kind of proposition it is. In particular, there is more than one way of establishing what the semantic contribution of the indexical phrase/apparently indexical phrase 'her evidence' is.

If the reconstruction that I put forward below indeed tracks the (somewhat implicit) way of thinking of the involved parties, then the resolution of what they take to be a disagreement about the possibility of rational akrasia depends either on what the target proposition more exactly is or, as I will explain in detail below, it simply depends on the resolution of a previous disagreement about how informative a subject's evidence is about what that subject's evidence is. Either way,

[3] In order to deal with 'basic' rational beliefs, we can assume that a total body of evidence that contains *p* supports *p* (although in a trivial sense), and that when one holds a 'basic' rational belief that *p* the proposition that *p* thereby counts as part of one's evidence.

progress on this issue would be made more likely by switching the debate to more basic questions, so to speak: either questions about the possible contents of higher-order beliefs of the relevant sort, or questions about the nature of evidence.

Here is how I will approach the issue. My point of departure will be the following question: can it be rational for a subject to believe that her evidence does not support *p* when her evidence does in fact support *p*? More generally, the question is whether it can be rational for one to form false beliefs about what one's evidence supports/fails to support. Call that question the 'rational fallibility question'.[4]

There are two proper answers to the rational fallibility question, 'yes' and 'no'.[5] Suppose there are cases where it is rational for a subject to believe that *her evidence does not support p*, even though her evidence in fact supports *p* (the answer is 'yes'). So the subject's evidence supports *p* while at the same time supporting the proposition that *her evidence does not support p*. The possibility that it is rational for a subject to believe that *p* while it is at the same time rational for her to believe that *her evidence does not support p*, then, is left open here. The only remaining question is whether other conditions for rational belief toward a proposition (other than evidential support) are satisfied for both the belief that *p* and the belief that *her evidence does not support p*.

Now suppose that, quite to the contrary, it cannot be rational for a subject to believe that *her evidence does not support p* when that subject's

[4] Relatedly, Whiting uses the label 'infallibilism' for the following view: if it is rational for a person to believe that it is irrational for them to believe that *p*, then it is irrational for them to believe that *p*—see section 7 of Daniel Whiting, "Higher-Order Evidence," *Analysis*, LXXX, 4 (2020): 789–807. That is, it cannot be rational for one to falsely believe that it is irrational for one to believe that *p*. Similarly, a negative answer to the rational fallibility question tells us that it cannot be rational for one to falsely believe that *one's total body of evidence does not support p*. The only kind of infallibilist view that I will be discussing here, however, is what Whiting calls a 'bottom-up' view: in cases of misleading higher-order evidence, the lower-order belief is rational while the higher-order one is not. I will not discuss what he calls 'top-down' infallibilism, according to which the higher-order belief is rational while the lower-order belief is not. (I do not think that top-downers can deal with cases where the subject's lower-order evidence *entails* that *p*, basically because entailment is a monotonic alethic relation, and no amount of higher-order evidence can "change" that—but I will not elaborate on this point here.)

[5] Those who answer 'yes' include Timothy Williamson, "Very Improbable Knowing," *Erkenntnis*, LXXIX, 5 (2014): 971–99; and Lasonen-Aarnio, "Enkrasia or Evidentialism?," *op. cit.* Those who answer 'no' include Michael Titelbaum, "Rationality's Fixed Point (or: In Defense of Right Reason)," in T. S. Gendler and John Hawthorne, eds., *Oxford Studies in Epistemology: Volume 5* (New York: Oxford University Press, 2015), pp. 253–94; and also Declan Smithies, "The Unity of Evidence and Coherence," in Nick Hughes, ed., *Essays on Epistemic Dilemmas* (New York: Oxford University Press, forthcoming). More on these authors' views below. For attempts to find a "middle way" between the opposite views here, see Kevin Dorst, "Evidence: A Guide for the Uncertain," *Philosophy and Phenomenological Research*, C, 3 (2020): 586–632; and David Christensen, "Akratic (Epistemic) Modesty," *Philosophical Studies*, CLXXVIII, 7 (2021): 2191–214.

evidence in fact supports *p* (the answer is 'no'). Then it cannot be rational for a subject to believe that *p* and at the same time rational for her to believe that *her evidence does not support p*. For it can only be rational for her to believe that *p* if her evidence supports *p*—in which case it is not rational for her to believe that *her evidence does not support p*. So a negative answer to the rational fallibility question immediately delivers the verdict that it is never rational for one to hold both of the attitudes that constitute the relevant kind of akratic state. The rational fallibility question, then, carries a lot of weight. A positive answer to it leaves the door open for the possibility of the relevant type of rational akrasia, whereas a negative answer to it rules that possibility out completely (without even making use of the observation that there would be some kind of incoherence involved in the realization of the akratic state).[6]

I. THE NEGATIVE ANSWER TO THE RATIONAL FALLIBILITY QUESTION

At first, the negative answer to the rational fallibility question might strike us as really odd. It says that it is never rational for a subject to believe that her evidence does not support *p* when her evidence supports *p*. But isn't it the case that, for pretty much any false proposition *f*, it is possible for it to be rational for one to believe that *f*?

That initial reaction notwithstanding, a defender of the negative answer could try and defend her view roughly as follows. It is not the case that, for *any* false proposition *f*, it can be rational for one to believe that *f*. For there is a class *C* of falsehoods such that no body of evidence ever confers any degree of support upon any of its members. The propositions that belong to *C* are not only false: it is also impossible for them to be true. Consider, for example, the propositions that *Maria is taller than herself*, or that *the brain is an organ and the brain is not an organ*, or that $5 = 7$. Arguably, no total body of evidence supports the truth of any of those propositions. A rough argument for that goes as follows: (i) supporting evidence is supposed to at least make that which it gives support to likely to be true, (ii) but nothing can make that which *cannot* be true likely to be true, (iii) the propositions in *C*

[6] For the claim that there are cases of the relevant sort (the answer to the rational fallibility question is 'yes'), even though one is still rationally required not to believe that *p* while at the same time believing that *one's evidence does not support p*, see Alex Worsnip, "The Conflict of Evidence and Coherence," *Philosophy and Phenomenological Research*, XCVI, 1 (2018): 3–44. That would be a requirement of *coherence*, however, which Worsnip takes to be an independent dimension of rationality (one that does not boil down to facts about what the evidence supports). This view contrasts with the one that has previously been put forward by Niko Kolodny, "How Does Coherence Matter?," *Proceedings of the Aristotelian Society*, CVII (2007): 229–63.

cannot be true, (iv) so none of the propositions that belong to C are ever supported by any body of evidence. Furthermore, evidential support is a necessary condition for rationality. So it cannot be rational for one to believe any of the propositions that belong to C. Finally, falsehoods about which bodies of evidence support which propositions also belong to C. Which total body of evidence supports which propositions is not a contingent matter.

The broadly Bayesian, probabilistic way of thinking about evidential support vindicates the idea that some falsehoods are never supported by any body of evidence (at a minimum, all logical falsehoods). According to this picture, whether a given body of evidence gives support to a proposition is a matter of how probable that proposition is conditional on that evidence, as measured by some probability function Pr.[7] Where $E = \{e_1, \ldots, e_n\}$ is a body of evidence, here construed as a set of propositions, we say that E supports p when $Pr(p \mid e_1 \wedge \cdots \wedge e_n) \geq t$. That means that the probability of p conditional on the evidence is equal to or above a certain threshold t. (We can make $t = 1$ if we want, and we can also have t vary with context—let us leave these options open.)

Conditional probabilities are computed through the ratio formula: $Pr(p \mid q) = Pr(p \wedge q)/Pr(q)$. The function Pr is distributed over a total set of possible worlds W, and we think of the probability of p as the probability of the set of possible worlds where p is the case.[8] Every member w of the total set of possible worlds W is assigned some initial probability $Pr(\{w\})$, so that $Pr(W) = 1$. The latter is actually one of the axioms of the probability calculus, from which it follows, together with the other axioms, that $Pr(\{\}) = 0$ (the empty set weighs nothing).[9] If f is true at no possible world, then, $Pr(f) = Pr(\{\}) = 0$ and, therefore, $Pr(f \mid q) = 0$, for any consistent proposition q. It follows that, for any consistent set $E = \{e_1, \ldots, e_n\}$, $Pr(f \mid e_1 \wedge \cdots \wedge e_n) = 0$. Reverting this back to where we started: no necessary falsehood is supported by any consistent body of evidence. Here, the relevant class C of falsehoods is the class of propositions that are false in all members of W.

[7] Pr might be interpreted as some kind of objective probability function. Alternatively, it might be taken to represent the credences of an ideally rational agent. For the former interpretation, see, for example, Timothy Williamson, *Knowledge and Its Limits* (Oxford: Oxford University Press, 2000); and, for the latter, see Anna-Maria Eder, "Evidential Probabilities and Credences," *The British Journal for the Philosophy of Science* (forthcoming).

[8] So here we use $Pr(p)$ as a shorthand for $Pr(\{w \in W : p \text{ is true in } w\})$. Similarly, $Pr(p \wedge q)$ is shorthand for $Pr(\{w \in W : p \text{ is true in } w\} \cap \{w \in W : q \text{ is true in } w\})$.

[9] The axioms of the probability calculus are laid out in the appendix below.

Of course, it is not that this theoretical framework must itself be part of the original idea that there is a class of necessary falsehoods such that no members of that class receive any support from any total body of evidence (and false claims about evidential support belong to that class—but see below for the alternative idea that a proposition about conditional probabilities has different truth-values in different possible worlds). That formal framework does indeed force us to be more explicit about our commitments, and it has the advantage of not making room for the possible ambiguities that pre-theoretical notions of evidential support make room for. But the defender of the original idea is not necessarily committed to using it. The ingredients of the negative answer to the rational fallibility question can be more minimal than that. Any combination of views of the following sort would do:

(a) Evidential support is an *alethic* relation between propositions. It is concerned with the bearing that certain propositions $e_1, \ldots, e_n$ have on the question of whether a given proposition p is true. In particular, where $E = \{e_1, \ldots, e_n\}$ is consistent and E entails that p, in the sense that it is *impossible* for all of $e_1, \ldots, e_n$ to be true and p to be false, it follows that E *does not* support $\neg p$ ($\neg p$ is the classical negation of p). How could E support $\neg p$ when it in fact entails that p? Different notions of impossibility might be deployed to explain what kind of entailment is relevant here, for example, conceptual impossibility, or *a priori* impossibility,[10] or even metaphysical impossibility.

(b) Evidential support is a necessary condition for the rationality of belief: it is rational for one to believe that p only if one's total evidence supports p. This is compatible with making further requirements for the rationality of belief, for example, that the subject appreciates or grasps the support relation between her evidence and p, or that she is able to reason from her evidence to p. It is just that the possibility that it is rational for one to believe that p when one's evidence does not support p is ruled out.

(c) False propositions about evidential support belong to a class C of falsehoods such that no body of evidence supports any member of that class. Which propositions more generally belong to the relevant class C? Any proposition q such that it is impossible for q to be true, in the same sense of 'impossible' that is at play in (a). Where q belongs to C, then, any total body of evidence entails that $\neg q$ in the sense of (a)—it is impossible for $e_1, \ldots, e_n$ to be true and $\neg q$ to be false, for any total body of evidence $E = \{e_1, \ldots, e_n\}$. So no total body of evidence gives support to q.

[10] For a notion of *a priori* entailment/possibility see David Chalmers, *Constructing the World* (New York: Oxford University Press, 2012), especially chapters 1 and 2.

We can see the combination of (a)–(c) as a *template* for negative answers to the rational fallibility question. Ideas about what more exactly the relevant class C of false propositions is, and concomitantly what the notion of entailment/impossibility in (a) is, can be taken from already existing views that forbid rationality or justification to believe falsehoods about rationality or justification. For example, both Declan Smithies and Michael Titelbaum put forward the idea that truths about rationality/justification are *a priori* knowable to us, so that their denials are *a priori* falsehoods (it is possible for one to know *a priori* that they are false).[11] We can apply that idea to the (a)–(c) template from above as follows: the target class C includes any false proposition q such that it is *a priori knowable* that $\neg q$, or q is *epistemically impossible*, and falsehoods about which bodies of evidence support which propositions fit that description.

Most importantly for my present purposes, however, it is crucial for this strategy or anything that resembles it that by 'it is not rational for the subject to believe that *her evidence does not support p*' it is meant that it is not rational for that subject to believe that *E does not support p* (assuming that subject's total evidence is actually E). It is *that* proposition that is not supported by any body of evidence, including E itself, so that it is not rational for the subject to believe it. As I will explain below, that proposition (a 'horizontal' proposition) is to be contrasted with another proposition: one that is not necessarily about E itself (a 'diagonal' proposition).

II. THE POSITIVE ANSWER TO THE RATIONAL FALLIBILITY QUESTION

Consider now the following kind of situation. A subject's evidence leaves open certain possibilities where her evidence differs from what it actually is. In the actual world, her evidence is E—but that very evidence does not exactly tell the subject that her evidence is E.[12]

With that in mind, we can now think of propositions that the subject might entertain such that (a) they are concerned with whether

[11] In Declan Smithies, "Moore's Paradox and the Accessibility of Justification," *Philosophy and Phenomenological Research*, LXXXV, 2 (2012): 273–300, Smithies uses the view that the relevant truths are *a priori* as part of a tentative explanation of why justification is accessible: one has/lacks justification to believe that p iff one has justification to believe that one has/lacks justification to believe that p. In Titelbaum "Rationality's Fixed Point," *op. cit.*, Titelbaum defends the Fixed Point Thesis, according to which mistakes about rationality are mistakes of rationality. His argument for that thesis actually features the claim that we are rationally required not to be akratic as a premise. I will get back to the Fixed Point Thesis in section V below.

[12] See Williamson, *Knowledge and Its Limits*, *op. cit.*, chapter 8, for an argument against the transparency of evidence.

that subject's evidence supports a proposition *p*, but (b) their truth-value changes from one possible world to the other, as the subject's own body of evidence changes from one possible world to the other. We might truly assert, then, that it is rational for that subject to believe that *her evidence does not support p*, even though her evidence is actually *E* and *E* supports *p*. Here the content encoded by the embedded sentence '*her evidence does not support p*' is not simply the proposition that *E does not support p*, but rather something like: *what 'her evidence' refers to does not support p* (see the next section for alternative formulations). The thought is that, since what 'her evidence' refers to might vary from one possible world to the other, it is not necessarily the case that the proposition expressed by 'her evidence does not support *p*' always has the same truth-value in all possible worlds.

The point can be put as follows. As soon as the embedded sentence '*her evidence does not support p*' expresses a proposition that *can* be true, despite the fact that it is actually false, there will be total bodies of evidence that confer support upon that proposition. (This idea is formalized in the appendix below.) After all, a total body of evidence can make that which can be true likely to be true, even though it cannot make that which cannot be true likely to be true. And so, as far as the condition of evidential support is concerned, it can be rational for the subject to believe that *her evidence does not support p* even though her evidence actually supports *p*. As far as the condition of evidential support for the rationality of a belief goes, then, we are free to answer 'yes' to the rational fallibility question.

This way of addressing the rational fallibility question vindicates some of our initial reactions to particular cases. Suppose that my evidence indicates to me that my evidence is a certain body of evidence *F* (or, perhaps, that the evidence that I have that bears on the issue of whether *p* is *F*). My evidence also supports the proposition that *F does not support p*. In such a case, it may be rational for me to believe both:

(1) *My evidence is F,*
(2) *F does not support p,*

in which case it also seems rational for me to believe:

(3) *My evidence does not support p.*

Since my evidence can support (1) and (2), it can also support (3). I believe that some of us would like to stick to that conclusion even under the further assumption that my evidence is actually *E* and that *E* supports *p* (say $F \subset E$: the propositions that belong to *F* are not enough to support *p*, but once we add the propositions that belong to

E but not to *F* we have a set that is enough to support *p*).[13] And the way of addressing the fallibility question sketched above allows us to do just that.

Contrast that to the negative answer to the fallibility question from the previous section. The line of reasoning there was: since my evidence is *E*, *E* supports *p*, and it is not a contingent matter that *E* supports *p*, my evidence does not (in fact cannot) give support to the proposition that *my evidence does not support p*—and so it is not rational for me to believe that proposition. These two lines of reasoning seem to issue contradictory verdicts, in that one of them says that it can be rational for me to believe that *my evidence does not support p* even when my evidence is *E* and *E* supports *p*, whereas the other one says that it cannot be rational for me to believe that *my evidence does not support p* in such a situation. But these two conclusions need not contradict each other: maybe what one of them is saying is that it can be rational for me to believe a proposition *x*, whereas the other one is saying that it cannot be rational for me to believe a proposition *y*—but *x* and *y* are not the same proposition (the fact that the same linguistic guise is used to express those two propositions notwithstanding). The next section says a bit more about what the relevant propositions could be.

III. THE CONTENT OF 'MY EVIDENCE DOES NOT SUPPORT *P*'

Suppose I sincerely utter: 'My evidence does not support *p*'. What proposition is thereby expressed in my context, or what is the content of my utterance? In what circumstances is what I said true?

Assume that, in the context where I utter that sentence—a certain time, place, speaker (that's me), and possible world[14]—my evidence is *E*. Assume furthermore that *E* supports *p*. Then we presumably want to say that *what I said* is false. At the very least, it must be false in the possible world where I made the utterance. (Maybe it can still be true in other possible worlds.) Similarly, we want to say that if I believe the proposition that was expressed through that utterance of mine, then I have a false belief.[15] And so, even if my evidence misleadingly suggests to me that my evidence is *F*, and *F* does not support *p*, we better not take the proposition that is expressed by the sentence I have uttered to be the proposition that *F does not support p*—because that is true, whereas what I said is false.

[13] See Lasonen-Aarnio, "Enkrasia or Evidentialism?," *op. cit.*, p. 611, for a similar setup.

[14] I am borrowing these elements of a context from David Kaplan, "Demonstratives," in J. Almog, J. Perry, and H. Wettstein, eds., *Themes from Kaplan* (New York: Oxford University Press, 1989), pp. 481–563.

[15] Consider the following abominable conjunction: the subject truly believes that *her evidence does not support p*, her evidence is *E*, and *E* supports *p*.

If we treat 'my evidence' as an indexical expression and give it a Kaplanian treatment, in such a way that expression lacks descriptive content,[16] then the proposition that is expressed by my utterance of 'My evidence does not support *p*' in the context described above is simply the proposition that *E does not support p*. The content of 'my evidence' is determined by a combination of my context of utterance and the *character* of that expression. (A character is a function from contexts to contents.) For example, the character of 'I' is something like: the speaker who is uttering the sentence. So we use a rule to the effect that 'I' refers to the speaker who is using that expression in order to determine what its content is. Similarly, the character of 'my evidence' is something like: the evidence that belongs to the speaker who is uttering the sentence. But the character by itself does less than specify the content of the indexical expression. When I say 'My evidence does not support *p*', then, the character of 'my evidence' teams up with the context of my utterance to determine the content *E*, in such a way that the proposition thereby expressed is again the proposition that *E supports p*. This interpretation accords with the alethic assessment from the previous paragraph: what I said when I uttered 'My evidence does not support *p*' is false. For we are assuming, remember, that the evidence that I have when I make the target utterance is *E*, and we are also assuming that *E* supports *p*.

Another proposition that I might be expressing when I utter 'My evidence does not support *p*', however, is what Robert Stalnaker calls a 'diagonal proposition'.[17] In the case at hand, that could be the proposition that *what 'my evidence' refers to at t is such that it does not give support to p*, or the proposition that *the sentence 'my evidence does not support p' (as uttered by me at t) expresses a truth*. These are just some candidate diagonal propositions—there are others still that are not about expressions at all. Consider, for example, something like: *S's evidence at t does not give support to p*, where '*S*' is a proper name that refers to me, and *t* is again the time of my utterance. In either case, the semantic contribution of the expression 'my evidence' to the content of the sentence I have uttered is not simply the body of evidence that constitutes my body of evidence in the original context of utterance: that expression now behaves like a description, instead of a directly referential term.

[16] For this way of interpreting indexical expressions, see John Perry, "The Problem of the Essential Indexical," *Noûs*, LXXXV (1979): 3–21; and, again, Kaplan, "Demonstratives," *op. cit.* Notice, however, that to say that the expression lacks descriptive *content* is not yet to say that it lacks descriptive *meaning*. (It has after all a *character*, which maps contexts into contents.)

[17] See Robert C. Stalnaker, "Assertion," *Syntax and Semantics*, LXXXV (1978): 315–32.

So the content of 'my evidence' (as used by me at *t*) is either *what 'my evidence' refers to at t* or *S's evidence at t*, or something along these lines. And since *what 'my evidence' refers to at t* or *S's evidence at t* need not be the same thing in all possible worlds, the proposition expressed by my original utterance will have different truth-values in different possible worlds.

The diagonal proposition is to be contrasted with the horizontal proposition. In the case at hand, the horizontal proposition is just the proposition that *E does not support p*, which we saw is expressed by my utterance of 'My evidence does not support *p*' in the context described above when 'my evidence' behaves as a sheer indexical. Let the world where I make that utterance be the actual world @. I make that utterance at a certain time *t* in @, and my evidence at *t* in @ is again *E*. Now let *w* be another possible world where my evidence at time *t* is *F*. *E* supports *p*, but *F* does not. Then, where the leftmost column contains the worlds that are part of the context of utterance at *t*, and the upper row contains the worlds where the target propositions are assessed as being true (T) or false (F), we can use the following matrix to illustrate the difference between the horizontal and the diagonal propositions:

	@	*w*
@	F	F
w	T	T

The diagonal proposition is the one that receives the evaluations marked in gray: it is the proposition that is false in @ but true in *w*. The proposition that is false in both @ and *w* is the (horizontal) proposition that *E does not support p*, and the one that is true in both @ and *w* is the (also horizontal) proposition that *F does not support p*. Now we can see that not only the horizontal, but also the diagonal proposition is false in @. So taking the content of my original utterance to be the diagonal proposition also allows us to say that *what I said* is false. For suppose that the content of my utterance of 'My evidence does not support *p*' in @ at *t* is the proposition that *what 'my evidence' refers to at t is such that it does not support p*. What 'my evidence' refers to in @ at *t* is *E*, and *E* supports *p*—so I said something that is actually false. But what I said might have been true. It is after all true in *w*.

IV. THE CONTENT OF 'IT IS/IS NOT RATIONAL FOR THE SUBJECT TO BELIEVE THAT HER EVIDENCE DOES NOT SUPPORT *P*'

I just made the contrast between the horizontal and the diagonal propositions by thinking of a situation where I myself utter the sentence 'My evidence does not support that *p*' in the actual world @ at

a certain time *t*, and where my evidence in that world at that time is *E*. My utterance of that sentence can express either of those propositions. But that was just a clarificatory expedient.

We can now think of two different attributors *A1* and *A2* assessing the epistemic situation of a subject *S* who is in a possible world *w* at a certain time *t*. *S*'s evidence in *w* at *t* is *E*. And *E* leaves open the possibility that *S*'s own evidence at *t* is *F*. We assume, again, that *E* supports *p*, whereas *F* does not. After reflecting on *S*'s situation, attributors *A1* and *A2* come up with the following assessments:

A1: It is not rational for *S* to believe that *her evidence does not support p*.
A2: It is rational for *S* to believe that *her evidence does not support p*.

A1 is accordingly a defender of the negative response to the rational fallibility question: it is not possible for it to be rational for *S* to believe that *her evidence does not support p* when the proposition that *her evidence does not support p* is false. Similarly, *A2* is a defender of the positive response to the rational fallibility question: it is rational for *S* to believe that *her evidence does not support p*, despite the fact that the proposition that *her evidence does not support p* is false in the possible world that *S* is in.

Notice that *S* herself does not have to utter anything here. The point is rather that *A1*'s utterance of 'It is not rational for *S* to believe that *her evidence does not support p*' is such that the embedded sentence '*her evidence does not support p*' stands for the proposition that *E does not support p*—a horizontal proposition. And *A2*'s utterance of 'It is rational for *S* to believe that *her evidence does not support p*' is such that the embedded sentence '*her evidence does not support p*' stands for the proposition that *what 'her evidence' refers to at t is such that it does not support p*, or the proposition that *S's evidence at t does not support p*—some such diagonal proposition. It turns out, then, that both *A1* and *A2* can be right at the same time: what one of them is saying does not in fact contradict what the other one is saying. And that can be so even if *A1* and *A2* agree that rationality is a function of evidential support, and they agree just as well that the denials of certain necessary truths are never supported by any body of evidence (suppose that there is no lingering difference in meaning behind their respective uses of 'rational' and 'supports').

To make the point explicit, *A1* and *A2* can both be right when they are assessing the situation of a subject *S* with the following features: *S*'s evidence in possible world *w* at time *t*, namely *E*, fails to give support to the proposition that *E does not support p*, even though it gives support to, say, the proposition that *S's evidence at t does not support p*.

A1 picks up on the former fact, whereas *A2* picks up on the latter one. Furthermore, when everything is made explicit in this way, *A2* might grant to *A1* that *E* does not give support to the proposition that *E does not support p*, since it is impossible for that proposition to be true and, therefore, it is false in every possible world that is left open by the subject's evidence *E*. And *A1* might grant to *A2* that *E* gives support to the proposition that *S's evidence at t does not give support to p*, since *E* might leave some possibilities open where *S*'s evidence at *t* is not the same as it is in *w*, and *S*'s evidence at *t* in those alternative scenarios does not give support to *p* (even though it does give support to *p* in *w*). The appendix shows how we can cash this out more precisely in a formal setting.

V. TWO EXAMPLES FROM THE LITERATURE

So *A1* and *A2* from above part ways—though they may not be aware of this—in that they interpret the embedded sentence 'her evidence does not support *p*' in different ways. *A1* says that it is not rational for *S* to have a false belief about what her evidence supports because *A1* thinks that it is not rational for *S* to believe the proposition that *E does not support p*. And *A2* says that it is rational for *S* to have a false belief about what her evidence supports because *A2* thinks that it is rational for *S* to believe the proposition that *S's evidence at t does not support p* (or something along these lines).

I do not take the situation between *A1* and *A2* to be merely possible or hypothetical. We can after all find defenders of the negative answer to the rational fallibility question in the literature who are most reasonably interpreted as thinking in the way that *A1* is thinking. And we can also find defenders of the positive answer to the rational fallibility question who are most reasonably interpreted as thinking in the way that *A2* is thinking.

As a representative example of the former, consider Titelbaum's Fixed Point Thesis, according to which, "No situation rationally permits an a priori false belief about which overall states are rationally permitted in which situations."[18] In the evidentialist version of that thesis, a subject's 'situation' will be her total body of evidence. To use Titelbaum's own notation, let $r(s)$ be the set of doxastic states that are rationally permissible for a subject who is in situation *s*. The Fixed Point Thesis says, then, that any false belief about the values of $r(s)$ is rationally forbidden, for any situation *s*. Where *o* is a doxastic state and it is false that $o \in r(s)$, no situation rationally permits a belief that

[18] Titelbaum, "Rationality's Fixed Point," *op. cit.*, p. 261.

$o \in r(s)$. It is *that* proposition that is never rational for one to believe. (In particular, it is not rational for a subject who is in the very situation s to believe it.) The proposition that $o \in r(s)$ is of course a horizontal proposition. Even though my situation might be different from what it actually is, s is the same situation in every possible world. Accordingly, even though the doxastic attitudes that are rationally permitted by my situation might be different from what they actually are, the same doxastic states belong to the set $r(s)$ in every possible world. That looks exactly like *A1*'s line of thought.

Of course, when it comes to the point that it is not rational for a subject to hold akratic combinations of attitudes, it matters that the target proposition is about the subject's situation *described as her situation*—a point that is explicitly acknowledged by Titelbaum.[19] To try and account for that, we can either include the self-identifying element in the guise or mode of presentation of the target proposition, or we can construe it as being true/false across a range of centered possible worlds (centered on a particular subject and/or time).[20] Either way, however, the relevant propositional content will still be $o \in r(s)$ which, if false, is false in every possible world, centered or not.

As a representative example of *A2*'s way of thinking, consider Lasonen-Aarnio's point that, in cases of introspection failure, one's evidence might give support to false propositions about what one's evidence supports/about what is likely to be true on one's evidence. She writes:

> Assume that p is part of Mosi's evidence. In virtue of containing p, Mosi's evidence makes q likely—but p is the only item of evidence Mosi has that supports q. Because q is likely, Mosi is permitted to believe q, or at least be highly confident in q. However, it is not certain on Mosi's evidence that p is part of his evidence. Perhaps, for instance, it is only 50% likely that p is part of Mosi's evidence. Then, it may only be 50% likely on Mosi's evidence that his evidence makes q likely and hence, that he is permitted to believe q.[21]

Let w be the possible world where E is Mosi's evidence at t, and $p \in E$, so that E supports q. Why would the observation that it is not certain on Mosi's evidence that p is part of his evidence in w at t be relevant to the question of whether it is rational for him to believe that *E supports q* (or that *q is likely conditional on E*)?

[19] *Ibid.*, p. 262.

[20] For a critical discussion about centered worlds, as well as further references to the literature on the problem of essential indexicals, see Shen-yi Liao, "What Are Centered Worlds?," *The Philosophical Quarterly*, LXII, 247 (2012): 294–316.

[21] Lasonen-Aarnio, "Enkrasia or Evidentialism," *op. cit.*, p. 611.

The answer, of course, is that it is not relevant to that question at all. But that observation is relevant to the question of whether it is rational for him to believe (in *w* at *t*) the proposition that *Mosi's evidence at t supports q*, since *Mosi's evidence at t* can be different things at different possible worlds. Or think of the proposition that *what 'his evidence' refers to at t is such that it supports q*. The observation that it is not certain on Mosi's evidence that *p* is part of his evidence in *w* at *t* is also relevant to the question of whether it is rational for Mosi to believe that proposition as well. Accordingly, Lasonen-Aarnio's embedded sentence 'his evidence makes *q* likely' in the last sentence of the quote from above does not just stand for the proposition that *E makes q likely*. It stands rather for a horizontal proposition that can be true in some possible worlds, false in others, as Mosi's evidence at *t* changes from one possible world to another. That looks exactly like *A2*'s line of thought.[22]

VI. DIFFERENT SOURCES OF DISAGREEMENT

If one epistemologist is saying that it is not rational for *S* to believe a proposition *p*, whereas another one is saying that it is rational for *S* to believe a different proposition *q*, where *p* and *q* are themselves not in conflict with each other, then why do those epistemologists act as if they disagree with each other? One reasonable explanation is that they disagree about *something*. But what could that be?

The considerations from above suggest one possibility: they might disagree about which proposition is 'the right one' when it comes to capturing the contents of what one's thoughts about what one's evidence supports would be, or about what it means to say 'to believe that *one's evidence does not support p*'. It might then strike us as a surprise that the resolution of the target epistemological issue depends on the resolution of questions concerned with mental content, propositions, and how they are connected to our ways of talking (philosophy of mind and language as a remedy for epistemological concerns). But another possibility is that the root disagreement is a disagreement about the nature of evidence, even when both epistemologists take evidence to be or to be reconstructible as sets of propositions. In this case, it would be a disagreement about how informative one's evidence is about what one's evidence is.

Suppose it is clear to *S* in *w* at *t* that her evidence is *E*. Then it becomes fixed throughout the whole space of possibilities that are left open by *E* that her evidence at *t* is *E*. There is no variation in what

[22] It also looks like Williamson's in his "Very Improbable Knowing," *op. cit.*

S's evidence at *t* could be anymore—at least not within the space of possibilities that are left open by *E* itself. In this case, the difference between assessing whether *S*'s evidence supports the proposition that *E does not support p* and assessing whether *S*'s evidence supports the proposition that *S's evidence at t does not support p* is less important. If one is supported by *E*, then so is the other.

And so there are epistemologists taking part in the debate about the possibility of rational akrasia who argue that evidence possession has a transparent status for the subject. In a sense, then, they are free to advance the view that it is not possible for it to be rational for one to have false beliefs about what one's evidence supports, regardless of whether those are horizontal or diagonal propositions.[23] The view that the subject's evidence is transparent to that very subject completely blocks the kind of situation whose alleged possibility leads to the positive answer to the rational fallibility question. Continuing with Lasonen-Aarnio's example from above, it could not be the case that *p* is part of Mosi's evidence even though Mosi's evidence does not give support to the proposition that *p is part of Mosi's evidence*. And the defender of the transparency of evidence can say this even though she takes 'Mosi's evidence' to have descriptive content. The meaning of that singular term is not just simply a set of propositions $\{a, b, \ldots, p\}$, and the proposition expressed by '*p is part of Mosi's evidence*' is not just the proposition that $p \in \{a, b, \ldots, p\}$.

Needless to say, that optimistic view about the transparency of evidence is controversial.[24] And, once we bracket that view, the difference between the horizontal and the diagonal proposition makes itself relevant to the rational fallibility question and the possibility of rational akrasia again. For whereas the proposition that $\{a, b, \ldots, p\}$ *supports q* will have the same truth-value across all possibilities, the truth-value of the proposition that *Mosi's evidence at t supports q* will vary across those possibilities. And, in particular, the truth-value of the latter might vary across the possibilities that are left open by Mosi's very evidence at *t*. Once we make room for the non-transparency of evidence, we thereby make room for that kind of situation.

VII. CONCLUDING REMARKS

Thus far in this paper I have rationally reconstructed negative and positive verdicts about the possibility of rational akrasia, respectively,

[23] For a representative, if rare, example of this line of thought, see Smithies, "The Unity of Evidence and Coherence," *op. cit.*

[24] See Williamson, *Knowledge and Its Limits*, *op. cit.*, chapter 8, for influential arguments to the contrary.

and thereby shown that epistemologists who issue those verdicts are not necessarily expressing mutually inconsistent propositions when they do so.

Certainly one could dispute the premises that, according to the reconstruction, lead to the negative and the positive answers to the rational fallibility question, and thereby to the question about the possibility of rational akrasia. For example, regarding the *negative* answer, one could dispute the thesis that no body of evidence gives support to the relevant kind of necessary falsehood. It may be, for example, that the testimony of a reliable logician gives the subject good evidence to believe a logical falsehood.[25] And, regarding the positive answer to that question, one could again dispute the presupposition that a subject might not know what her own evidence is. But none of that means that the target epistemological views have not been rationally reconstructed. For the reconstruction is not supposed to imply that the reasoning that leads to each of those views is undisputable. It is just that the reasoning is good (because valid) and it departs from premises that seem to be true or are *prima facie* plausible.

The reconstruction offered here seems to be neutral on the debate at issue, because it vindicates responding 'It depends' to the question 'Is rational akrasia possible?' Let an ascription of rational akrasia be a sentence of the form 'It is rational for the subject to believe that *p*, and it is also rational for her to believe that *her evidence does not support p*'. If the fragment '*her evidence does not support p*' stands for a horizontal proposition, the reconstruction vindicates answering 'Yes', but if that fragment stands for a diagonal proposition, the reconstruction vindicates answering 'No' to the question about the possibility of rational akrasia. On the face of it, this reconstruction seems to discourage us from following up with: 'But is rational akrasia after all possible or not?' The answer is always going to be 'It depends', or 'It depends on what you mean by "*her evidence does not support p*"'.

As pointed out by an anonymous referee for this JOURNAL, however, the rational reconstruction offered above also invites us to consider the following question: would believing that *p* while at the same time believing that *my evidence does not support p* make me akratic/incoherent on the assumption that the fragment '*my evidence does not support p*' expresses a diagonal proposition such as the ones mentioned above? It seems like it does.

Let '*S*' again be a proper name that refers to me, and assume that I know that *I am S*. It then seems incoherent for me to believe that

[25] This venue of dispute/disagreement was suggested to me by an anonymous referee for this JOURNAL, to whom I am very thankful.

p at *t* while at the same time believing that *S's evidence at t does not give support to p*. (Here we stop using the personal pronoun to make clear exactly what the target proposition is.) Of course, if I were to believe that *S is some person other than myself*, then there would be no incoherence—but I actually believe that *I am S*. Finally, notice that it is perfectly possible for a total body of evidence to give support to *p* and to the proposition that *S's evidence at t does not give support to p*. After all, the latter proposition can be true in some possible worlds. And nothing seems to get in the way of the possibility that such a total body of evidence is my own total body of evidence at time *t* (nothing other than the implausible assumption, again, that my evidence always contains all the accurate information about what is part of my total body of evidence).

A defender of the possibility of rational akrasia could say all this while granting to her opponent, *arguendo*, that the horizontal proposition that is expressed by '*my evidence does not support p*'—something of the form *E does not support p*—is never supported by any body of evidence. But she will keep pressing on the point that it is incoherent for me to believe that *p* at *t* while at the same time believing that *S's evidence at t does not give support to p* (given that I also believe that *I am S*). It seems, then, that the reconstruction offered above is not neutral on the debate about the possibility of rational akrasia after all. On closer inspection, it suggests that the defender of that possibility has the upper hand at the end of the day. So the reconstruction does not issue the verdict that we have to answer 'It depends' in response to the question 'Is rational akrasia possible?'. On the assumption that believing that *p* while believing the aforementioned diagonal proposition makes me akratic, the answer is rather 'Yes'.

APPENDIX

Here I formalize the two lines of thought described in the paper. One of them, remember, says that *S*'s evidence in *w* at *t*, namely *E*, does not give support to the proposition that *E does not support that p* (assuming that *E* in fact gives support to *p*), whereas the other one says that *S*'s evidence in *w* at *t* supports the proposition that *S's evidence at t does not support p* (or something along these lines, as long as there is change of reference from one world to the next). For that reason, the defender of the former line of thought (*A1*) utters 'It is not rational for *S* to believe that *her evidence does not support p*', whereas the defender of the latter line of thought (*A2*) utters 'It is rational for *S* to believe that *her evidence does not support p*'.

To formalize this, I use the framework of possible worlds semantics. To represent the situation of a subject *S* across a relevant range of

possibilities, we use a frame $\langle W, R, Pr \rangle$, where W is a total set of possible worlds or ways the world might be, R is an accessibility relation that maps each member of W into a subset of W, and Pr is a prior probability distribution over W. We can assume that the members of W are centered on a time and subject, so that every $x \in W$ contains reference to time t and subject S (which I will leave implicit in order to avoid cluttering).

Propositions are construed as sets of possible worlds. A proposition p is true in a possible world x when $x \in p$. $R(x)$ is the set of possible worlds that are left open by the subject's evidence in x—or the ways the world might be according to the subject's evidence in x. A proposition p counts as part of the subject's evidence in x, then, when $R(x) \subseteq p$.

The set $R(x)$ thereby encodes everything that is part of the subject's total body of evidence in x. A prior probability is assigned to every proposition that is constructible from W as follows: where $p = \{x_1, \ldots, x_n\}$, $Pr(p) = Pr(\{x_1\}) + \cdots + Pr(\{x_n\})$. What makes this possible is the fact that Pr 'starts off' by assigning a real number between 0 and 1 to every member of W. Since Pr is a probability function, it abides by the following axioms:

(P1) $Pr(W) = 1$,
(P2) $0 \leq Pr(p) \leq 1$, for all $p \subseteq W$,
(P3) $Pr(p \cup q) = Pr(p) + Pr(q)$ for all $p, q \subseteq W$ such that $p \cap q = \{\}$.

Next, we explicate the notion of evidential probability. $Pr_x(p)$ is the evidential probability of p in x, which is just the probability of p conditional on the evidence that is possessed by the subject in x. The evidential probability of p in w is computed from the prior probability distribution as follows: $Pr_x(p) = Pr(p \mid R(x)) = Pr(p \cap R(x))/Pr(R(x))$. For modeling purposes, we assume that $Pr(\{x\}) > 0$ for all $x \in W$, so that these conditional probabilities are always defined.

With all this in place, we can now establish some common ground between *A1* and *A2* regarding evidence possession and evidential support. They agree that $R(w)$ encodes all the evidence E that is possessed by S in w. That is, E is a set of propositions such that $p \in E$ if and only if $R(w) \subseteq p$. They also agree that evidential support is a matter of high (possibly maximal) evidential probability. More formally, they agree that for any body of evidence F and $x \in W$, assuming that F is the set of propositions such that $p \in F$ if and only if $R(x) \subseteq p$, F gives support to p if and only if $Pr_x(p) = Pr(p \mid R(x)) \geq t$, where $0.5 < t \leq 1$.

A1 and *A2* are again looking at *S*'s situation in w (at a time t). Both of them make their assessments about *S*'s situation by uttering a sentence containing the fragment '*her evidence does not support p*'. But they take that expression to stand for different propositions. Specifically, *A1* takes it to stand for the following proposition:

$$a_1 = \{x \in W : Pr_w(p) < t\} = \{x \in W : Pr(p \mid R(w)) < t\}.$$

More informally, a_1 is the set of worlds where *E* does not support *p*. We can also represent it thus: $\{x \in W : E$ does not support p in $w\}$. Notice that 'in w' is *otiose* there (in contrast, see a_2 below). That is precisely because the proposition that *E does not support p* has the same truth-value in every possible world. Formally, this means that $Pr(p \mid R(w)) < t$ has the same truth-value in every possible world. Where '$\models$' is the truth-making relation, or truth in the world of a model $M = \langle W, R, Pr\rangle$:

$$M, x \models Pr(p \mid R(w)) < t \text{ iff } Pr(p \cap R(w))/Pr(R(w)) < t, \text{ for any } x \in W.$$

So either a_1 is the whole set W or it is the empty set $\{\}$. But we were assuming that *E* supports *p*, in such a way that:

$$M, x \models Pr(p \mid R(w)) \geq t, \text{ for any } x \in W.$$

So a_1 is the empty set after all, that is, $a_1 = \{\}$ (a_1 is true in no possible world). And since:

$$M, x \models Pr(\{\} \mid R(x)) = 0, \text{ for any } x \in M,$$

it follows that a_1 is not supported by any body of evidence. In particular, $Pr(a_1 \mid R(w)) = 0$, so that *S*'s evidence in w does not give support to a_1. What does *A1* conclude from this? She declares that 'It is not rational for *S* to believe that *her evidence does not support p*', because *S*'s evidence in w does not give support to a_1, and a_1 is the proposition that *A1* takes the embedded sentence '*her evidence does not support p*' to stand for.

In contrast, *A2* takes that embedded sentence to stand for the following proposition:

$$a_2 = \{x \in W : Pr_x(p) < t\} = \{x \in W : Pr(p \mid R(x)) < t\}.$$

More informally, a_2 is the set that contains every possible world x such that *S*'s evidence there, in x, does not support *p*. We can also represent it thus: $\{x \in W : S$'s evidence in x does not support $p\}$. Now notice that 'in x' (or '$R(x)$') is bound in the definitional notation of this set—now *there is* a connection between what goes into the set and what

defines what goes into the set. In contrast to a_1, a_2 can have different truth-values at different worlds. So even though, as before:

$$M, w \models Pr(p \mid R(w)) \geq t,$$

there still can be worlds $x \in W$ such that:

$$M, x \models Pr(p \mid R(x)) < t,$$

and $R(w)$ contains some of these worlds x. Besides, the probabilistic distribution might also be such that $Pr(a_2 \cap R(w))/Pr(R(w)) \geq t$, in which case $Pr(a_2 \mid R(w)) \geq t$. *A2* will then declare that 'It is rational for *S* to believe that *her evidence does not support p*', because *S*'s evidence in w gives support to a_2, and a_2 is the proposition that *A2* takes the embedded sentence '*her evidence does not support p*' to stand for.

Since both of these can hold at the same time:

$$M, w \models Pr(a_1 \mid R(w)) < t,$$
$$M, w \models Pr(a_2 \mid R(w)) \geq t,$$

A1 and *A2* can both be right at the same time (given *what they mean* through their utterances).

LUIS ROSA

University of Cologne