Episteme, 14, 3 (2017) 263–274 © Cambridge University Press
doi:10.1017/epi.2017.24

AMBIGUOUS RATIONALITY

TIMOTHY WILLIAMSON
timothy.williamson@philosophy.ox.ac.uk

ABSTRACT

The paper distinguishes a content-oriented conception of rational belief, which concerns support relations between the proposition believed and one's evidence, from a disposition-oriented conception of rational belief, which concerns whether someone generally disposed to conform their belief to their evidence would believe the given proposition in the given circumstances. Neither type of rationality entails the other. It is argued that conflating the two ways of thinking about rational belief has had damaging effects in epistemology.

INTRODUCTION

As a step towards making up one's mind, one may ask what it is *rational* to believe, or to do. One may ask the question about oneself or about others. This paper focuses on the case of belief, but its argument may well generalize to the case of action. I will argue that our thinking about what it is rational to believe is infected by equivocation, with damaging consequences for epistemology.

1. TWO CONCEPTIONS OF WHAT IT IS RATIONAL TO BELIEVE

What is it rational for me, in my current circumstances, to believe? The question can be understood in more than one kind of way.

This paper concerns only the *epistemic* rationality of belief, not its *pragmatic* rationality. For instance, it may be pragmatically rational for each runner in a race to believe that he is the best, simply because believing it increases his probability of winning, though in no case (we may suppose) to as much as 50%. It may also be pragmatically rational for a mathematician to believe that she will eventually solve a problem, simply because believing it increases her probability of solving it, though (we may suppose) not to as much as 50%. This paper does not concern that kind of consequentialist rationality of belief, whether the consequences are evaluated in terms of truth-related or truth-unrelated goods. Rather, it concerns more directly epistemic evaluations of belief. But even with that proviso, the question 'What is it rational to believe?' can be understood in more than one non-consequentialist kind of epistemic way.

One might understand the question as asking which contents of potential belief are suitably related to the content of one's current evidence. This is the idea:

Content-oriented schema
It is rational$_{cont}$ to believe p if and only if one's evidence supports p.

Here both 'evidence' and 'supports' are schematic terms, to be filled in according to one's theoretical proclivities. For instance, I have argued elsewhere that one's evidence is the totality of what one knows (Williamson 2000), while some other epistemologists equate it with something more phenomenal. A strong relation of support is entailment (the evidence entails p); a weak relation of support is consistency (the evidence does not entail $\neg p$); there are intermediate relations of support, including probabilistic ones (the probability of p conditional on the evidence is high). For now, I will be neutral between these alternative fillings-in of the schema.

The content-oriented schema is 'propositional' rather than 'doxastic'. Even when their evidence supports p, someone may believe p without doing so *because* their evidence supports p. They may come to believe p in some different and foolish way, while ignoring the genuinely relevant features of their evidence. In those circumstances it is still rational$_{cont}$ for them to believe p, though believing p in that way is irrational in some other sense.

We apply the term 'rational' to agents as well as to acts (including the acquisition and maintenance of beliefs). There is a two-way interaction between the two sorts of application. We judge the rationality of agents by the rationality of their acts, but we also judge the rationality of acts by whether they are what a rational agent would do. Such judgments involve some degree of idealization. In some circumstances, perhaps, any normal human would act irrationally. Thus the relevant rational agent is more rational than any normal human. But the idealization should not be taken too far. Presumably, it would not have been rational for mathematicians in 1900 to believe Fermat's Last Theorem, since they had no proof, even though a *perfectly* rational mathematician in 1900 would have instantly constructed a proof. In judging the rationality of an act by whether a rational agent would have done it, we typically impose a medium level of idealization. Such a medium level of idealization is in common use. For instance, in trying to determine the morally right thing to do, we may consider what a good person would do in the circumstances. We are not asking what a normal human would do, but we are also not asking what a perfectly good divinity would do. The good person we envisage is somewhere in between.

Judging the rationality or goodness of an act by what a rational or good person would do in the circumstances may look unnecessarily indirect. Why not judge the rationality or goodness of the act directly? But asking what a rational or good person would do may be an effective way of bringing the cognitive power of one's imagination to bear on the issue (compare Williamson 2016). Our imaginative grip on what people would do may be firmer than our grip on what abstract standards of morality or rationality demand. For instance, suppose that your prototype of the good person is Nelson Mandela. Then you might work out the right thing to do by imagining what Nelson Mandela would do in your circumstances. That is one main function of role models. We can also use fictional characters for the same purpose. If your prototype of the good person is Elinor Dashwood in *Sense and Sensibility*, you might work out the right thing to do by imagining what Elinor Dashwood would do in your circumstances. We may also use role models in judging rationality as well as in judging goodness: 'How would so-and-so approach this problem?' Imagining what the good or rational person would do in your circumstances is just a more abstract and idealizing version of the same process.

Talk of what a good or rational person would do in counterfactual circumstances requires a further clarification. Do 'good' and 'rational' characterize the person as they actually are, or as they would be in the counterfactual circumstances? The latter reading is the relevant one. For if someone is actually a good or rational person, and would do something A in counterfactual circumstances in which they would *not* be a good or rational person, that does not seem to go any way towards showing that in those circumstances A would be a good or rational thing to do. For instance, if someone is rational when sober and irrational when drunk, and is now sober, what he would do when drunk (say, provoking a fight) does not show what it would be rational for him to do when drunk (say, going home to bed). Conversely, if someone is actually not a good or rational person, but would do A in counterfactual circumstances in which they *would* be a good or rational person, that does seem to go some way towards showing that in those counterfactual circumstances A would be a good or rational thing to do. Thus, when we speak of what a good or rational person would do in counterfactual circumstances, 'good' or 'rational' belongs in the consequent of the counterfactual conditional, not in its antecedent: if those circumstances were to obtain, it would be that good or rational people did the thing in question. If we take what Nelson Mandela would do in counterfactual circumstances as a guide to what it would be right to do in those circumstances, we do so because we assume that he would still be a good person in those circumstances.

In the epistemic order, the goodness or rationality of agents sometimes precedes the goodness or rationality of acts. But in the metaphysical order, one might expect, the goodness or rationality of acts always precedes the goodness or rationality of agents. For rationality with respect to belief, the natural proposal is that to be a rational agent is to have a general *disposition* to believe just what it is rational$_{\text{cont}}$ to believe. Philosophers have tried to reduce dispositions to counterfactual conditionals, and that might be attempted here too. I am sceptical about such reductions, but will remain neutral about them for present purposes.

Filling in the content-oriented schema involves giving a more detailed content to 'what it is rational$_{\text{cont}}$ to believe', along the lines of 'what one's evidence supports'. Thus, at least at a first pass, to be a rational agent in a corresponding sense is to have a general disposition to believe just what one's evidence supports.

Some tweaking of that definition might be required for highly permissive conceptions of rationality$_{\text{cont}}$, for instance when evidential support is mere consistency with the evidence, since then one's evidence may simultaneously support p and support $\neg p$, although a rational person will presumably not believe both (however, it is not plausible that when a fair coin is about to be tossed, it is rational to believe that it will come up heads, and rational to believe that it will come up tails).[1] It is more or less of an idealization to suppose that all rational people would believe exactly alike in the same circumstances with the same evidence. Nevertheless, for the sake of simplicity and clarity, this paper will not fuss about violations of that idealized assumption.

1 'Belief' here means flat-out belief, not just high credence. One can have extremely high credence that one's ticket will not win the lottery (as measured by one's betting behaviour, for instance), without believing flat-out that it will not win; one does not throw the ticket away. No good norm of rationality forbids high credence without flat-out belief in each member of an inconsistent set of propositions.

https://doi.org/10.1017/epi.2017.24 Published online by Cambridge University Press

Another issue for the first-pass definition of 'rational person' is that believing whatever one's evidence supports involves cluttering up one's mind with a host of well-supported beliefs irrelevant to one's interests. To finesse these issues, let us say at a second pass that to be a rational person is to have a general disposition to *conform* one's beliefs to what it is rational$_{\text{cont}}$ to believe, that is, to conform one's beliefs to what one's evidence supports, where 'conform' is itself a schematic term to be filled in according to the theorist's predilections.

Of course, one can be disposed to do something without *always* doing it. A fragile vase is disposed to break when struck, but it does not always break when struck; for instance, when it is protectively bubble-wrapped. Similarly, although a rational person is disposed to conform their beliefs to what their evidence supports, it does not follow that they always conform their beliefs to what their evidence supports. In unfavourable circumstances, they may fail to do so.[2] To require unrestricted infallibility would be to impose a standard more suitable for gods than for humans.

As we have in effect seen, the idea of a rational agent naturally gives rise to a second kind of way to understand the original question 'What is it rational to believe?'. For one can ask what a rational agent *would believe* in the same circumstances with the same evidence. By unpacking 'rational agent' as 'agent disposed to conform their beliefs to what their evidence supports', we make the two understandings of the question maximally comparable with each other (which will turn out to be a concession to my opponents). On the second understanding, the question is what someone disposed to conform their beliefs to what their evidence supports would believe in these circumstances with this evidence. Thus we have a second schema:

Disposition-oriented schema
It is rational$_{\text{disp}}$ to believe p if and only if in the same circumstances with the same evidence someone disposed to conform their beliefs to what their evidence supports would believe p.

Like the content-oriented schema, the disposition-oriented schema is propositional rather than doxastic. Even if in the same circumstances with the same evidence someone disposed to conform their beliefs to what their evidence supports would believe p, someone else may believe p without doing so *because* their evidence supports p, or *because* they are disposed to conform their beliefs to what their evidence supports. They may come to believe p in some different and foolish way, while ignoring the genuinely relevant features of their evidence. In those circumstances with that evidence it is still rational$_{\text{disp}}$ to believe p, though believing p in that way is irrational in some other sense.

For clarity and simplicity, let us hold the interpretations (whatever they are) of the shared schematic terms 'evidence' and 'supports' fixed between the two schemas. We do not have to choose one schema over the other; we could use the two subscripted terms 'rational$_{\text{cont}}$' and 'rational$_{\text{disp}}$' in tandem, without treating them as synonymous.

Unfortunately, there is a tendency in the epistemological literature and elsewhere to use phrases like 'it is rational to believe p' as though they were governed by both schemas simultaneously, evaluating their truth-value sometimes according to what the evidence supports, sometimes according to what a rational person would believe in those

2 See Martin (1994), Lewis (1997), and Bird (1998).

circumstances with that evidence, and combining the results. On the face of it, that is just to equivocate between what it is rational$_{cont}$ to believe and what it is rational$_{disp}$ to believe, with all the consequent danger of smoothing the way for fallacies. It smuggles in the assumption that it is rational$_{cont}$ to believe something if and only if it is rational$_{disp}$ to believe it, in other words, this principle:

Equivalence Schema
One's evidence supports *p* if and only if in the same circumstances with the same evidence someone disposed to conform their beliefs to what their evidence supports would believe *p*.

Note that the word 'rational' does not occur in the equivalence schema, so no interpretation of 'rational' by itself can vindicate the schema. Rather, the equivocation on 'rational' smuggles in a substantive assumption about evidence and support. Even if that assumption is in fact correct, we should accept it only after checking that it passes theoretical muster, rather than letting it sneak in under our radar.

By 'equivocation' here, I do not mean the sort of ambiguity usually recorded in a plurality of lexical entries for the same word, or homophonic words. The point is not that there are two separate practices of using 'rational', one associated with the content-oriented schema, the other with the disposition-oriented schema. If the practices were kept properly separate, the ambiguity would be harmless, as with 'bank', for there would be no commitment to the equivalence schema. Rather the trouble is that both schemas are associated with the *same* practice of using the word 'rational'. Within that single practice, we shift between applying the term according to the content-oriented standard and applying it according to the disposition oriented-standard, combining the results and thereby incurring commitment to the equivalence schema. That is why the equivocation is dangerous.

For similar reasons, the equivocation is not to be understood as mere dependence of the extension of the word 'rational' on the conversational context, with the content-orient schema determining its extension in some contexts and the disposition-oriented schema determining its extension in others. By itself, that too would be harmless. The problem is that shifting from one standard to another is not treated as a relevant change of context, so the dangerous agglomeration is not blocked.

That our practice commits us to the equivalence schema does *not* mean that when we read it, we find it compelling or even plausible. As well-trained philosophers, we are immediately on the look-out for counterexamples. The point is just that our practice of using 'rational' is in good epistemic order only if the equivalence schema is correct. In general, explicit articulations of our implicit commitments often make us notice the problematic nature of those commitments; in extreme cases, by making their inconsistency manifest. The practice may depend on *not* explicitly articulating its commitments.

Of course, if the equivalence schema were correct, our shifts between one standard and another would be comparatively harmless, although they would still involve taking a substantial theoretical principle for granted without critical reflection. But is the equivalence schema correct? One might try arguing for it thus:

The simple-minded argument
Suppose that one is in circumstances C with evidence E. Let S be a subject in C with evidence E disposed to conform their beliefs to what their evidence supports. Then, in C, S conforms S's

https://doi.org/10.1017/epi.2017.24 Published online by Cambridge University Press

beliefs to what E supports, so S believes *p* if and only if E supports *p*. (i) Suppose that E supports *p*. Hence, in C, S believes *p*. Therefore, in the same circumstances with the same evidence, someone disposed to conform their beliefs to what their evidence supports would believe *p*. (ii) Conversely, suppose that, in C, S believes *p*. Then E supports *p*. QED.

Several things are wrong with the simple-minded argument. Most notably, it assumes that a disposition to do something will make one do that thing in any given circumstances. As we have already seen, that assumption is problematic. Thus the simple-minded argument fails.

It is one thing to rebut an argument for a conclusion, another to rebut the conclusion itself. The latter task is the business of the next section.

2. THE NON-EQUIVALENCE OF THE TWO CONCEPTIONS

I will argue that the equivalence schema fails in both directions, on any reasonable way of filling in its schematic terms. In order to do so, I will use a non-standard type of sceptical scenario that I introduced elsewhere for related reasons (Williamson Forthcoming). I use the sceptical scenario for non-sceptical purposes.

A background assumption of the argument is that our evidence often does support propositions in standard ways; it is not rational to be a Pyrrhonist sceptic, disposed to suspend belief about everything. Like the rest of us, the rational person would have false beliefs in a sceptical scenario. Non-sceptics may find little to admire in the Pyrrhonist's self-imposed ignorance, for instance when that ignorance concerns the needs of others.

Imagine a special device, the *brain-scrambler*, which emits waves of some sort with a selective scrambling effect on the brains of those at whom it is pointed. The waves inflict no permanent damage, and do not even change what programme it would be natural to describe the brain as running, but they occasionally alter the contents of unconscious short-term working memory, so that some computations produce incorrect results. The waves do not affect memory in other ways. Under the misleading influence of the brain scrambler, a normal subject may confidently announce that $17 + 29 = 33$. Similarly, consider Innocent, a normal rational agent and excellent mathematician who sincerely and confidently announces that 179 is and is not prime, because a scrambled piece of reasoning yields that conclusion, and a scrambled application of a contradiction-detector failed to sound the alarm in retrospect. Innocent has not gone temporarily mad. Rather, she is like someone doing a long calculation on paper with a prankster standing behind her, who from time to time when she is not looking jumps out, erases some of her figures, replaces them with other ones, and jumps back without her noticing. The brain-scrambler has the advantage of interfering with her memory to prevent her from noticing the changes. Innocent's attitude to the proposition that 179 is and is not prime arguably amounts to belief: she acts on it, for example when her career is riding on it in a mathematics test. We may assume that these effects are deterministic: any other competent calculator embarking on that very calculation would make the same errors when the scrambler was turned on in the same situation. Thus, in the same circumstances with the same evidence, anyone disposed to conform their beliefs to what their evidence supports would believe that 179 is and is not prime. But Innocent's evidence does not support that contradiction. Indeed, whatever that evidence is, it is inconsistent with the contradiction, for the

contradiction is inconsistent with itself. Furthermore, the contradiction has probability zero on the evidence, whatever it is, by the axioms of mathematical probability theory. Thus the equivalence schema fails in the right-to-left direction.

For similar reasons, the equivalence schema fails in the left-to-right direction. The brain-scrambler may cause Innocent to refuse to believe a tautology, even though her evidence entails it (because everything does) and it has probability 1 on her evidence.

The underlying structure of the argument does not really depend on the special cases of belief in a contradiction and non-belief in a tautology. We can just as easily use a case where the brain scrambler makes Innocent believe an ordinary contingent proposition inconsistent with her evidence, and fail to believe an ordinary contingent proposition entailed by her evidence. For instance, they may be propositions about the qualifications of some candidates for a job. Clearly, she can act on such beliefs. They may determine who gets the job. Crucially, the brain scrambler does *not* obliterate Innocent's previous evidence about the candidates' qualifications; it merely causes her to make wildly fallacious inferences from that evidence.

The key point is that in normal circumstances, when the scrambler is switched off, Innocent has the general dispositions of a rational person, and switching on the scrambler does not change her general dispositions. Its effect is to interfere temporarily with the operation of her general dispositions, not to destroy them, just as a fragile vase remains fragile even when it has been wrapped in protective material. Nothing more than stopping the interference (switching off the scrambler, removing the wrapping) is needed to enable the disposition to manifest again, unlike cases where interference temporarily makes something lose a disposition. Thus Innocent remains a rational agent even while the brain scrambler is switched on and pointed at her, just as one can remain a rational agent even while the prankster is interfering with one's calculations. In coming to believe that 179 is and is not prime, Innocent does what someone with the disposition to conform their beliefs to what the evidence supports would do in her circumstances, which include the scrambler interfering in its predetermined characteristic way.[3]

3. WHAT HARM DOES CONFLATING THE TWO CONCEPTIONS DO?

To conflate rationality$_{\text{cont}}$ and rationality$_{\text{disp}}$ is to commit oneself in effect to the invalid equivalence schema, and to confuse behaviour with character. It also has more specific epistemological consequences. In particular, it will tend to warp theorizing about the nature of evidence.

Consider a good case, in which one is disposed to conform one's beliefs to what one's evidence supports, and one truly believes in the usual way that one has hands, and a corresponding bad case, a sceptical scenario in which one is a brain in a vat with the same dispositions as the brain in the good case, and one falsely believes in a similar way that

3 For a closely related distinction see Lasonen-Aarnio (2010). She uses a norm of something analogous to rationality$_{\text{disp}}$ to explain away nicely some supposed cases of knowledge defeat: in effect, she points out that a rash subject may do something analogous to believing what it is rational$_{\text{cont}}$ but not rational$_{\text{disp}}$ to believe. Compare also the derivation of secondary evidential norms for assertion from the primary norm of knowledge (Williamson 2000: 257; for a recent application see Benton (2013: 357–8), drawing on DeRose (2002: 180; 2009: 94–5)).

https://doi.org/10.1017/epi.2017.24 Published online by Cambridge University Press

one has hands. As the cases have been set up, one's dispositions do not vary between the cases; it is just that circumstances are helpful in the good case and unhelpful in the bad case. In both cases, one is disposed to conform one's beliefs to what one's evidence supports, and so believes just what someone so disposed would believe in those circumstances with that evidence. Thus, in each case, one believes just what it is rational$_{\text{disp}}$ to believe in that case. Therefore, given the equivalence schema, in each case, one believes just what it is rational$_{\text{cont}}$ to believe in that case. But, we may assume, what one believes in the bad case is just what one believes in the good case (that assumption ignores semantic externalism about the contents of one's belief, which we may do for the sake of argument, since conceding it would only make matters worse for my opponents).[4] Consequently, what it is rational$_{\text{cont}}$ to believe in the bad case is just what it is rational$_{\text{cont}}$ to believe in the good case. In other words, what one's evidence supports in the bad case is just what it supports in the good case. This makes it hard to avoid the conclusion that one has exactly the same evidence in the good and bad cases. For suppose that one's evidence differs between the two cases — naturally, in ways undetectable from within the bad case. Then the degree to which one's evidence supports some propositions will vary between the two cases. For instance, suppose that the proposition *e* is part of one's evidence in the good case but not in the bad case. Thus one's evidence entails *e* in the good case but not in the bad case (although even in the bad case one's evidence may make *e* probable). Given such variations in degree of support, it is virtually inevitable that there will be propositions *p* supported just enough for rational belief in one case (say, the good one) but not enough for rational belief in the other case (say, the bad one). That contradicts the conclusion reached above from the equivalence schema, that one's evidence supports exactly the same conclusions in the two cases. Thus we are led to deny the supposition that one's evidence differs between the two cases. We end up asserting that one has the same evidence in the good and bad cases. That, I have argued elsewhere, is a disastrous conclusion (Williamson 2000: 173–81). Whether one accepts or rejects that verdict, it should be clear that the 'same evidence' claim is a highly contentious theoretical judgment, for which strong arguments would be needed. One should not allow oneself to be manipulated into accepting a highly contentious picture of evidence just through conflating two fundamentally distinct conceptions of rationality, as articulated in the equivalence schema.

It is less easy to see that the equivalence schema has false instances in the usual sceptical scenarios like that of the brain in the vat than in sceptical scenarios like that of the scrambled brain.[5] But the result could be that we are more liable to distort our epistemology to accommodate the schema in the former cases than we are in the latter.

What is clear enough is that it is rational$_{\text{disp}}$ for the brain in the vat to believe that it has hands. Given the unsoundness of the equivalence schema, we should not jump to the conclusion that it is rational$_{\text{cont}}$ for the brain to believe that it has hands. For if the brain in the

4 This brain was envatted only the night before, so semantic externalism allows it to have beliefs about hands, not just about patterns of neural activation. For the sake of argument, we may ignore the differences in belief flowing from the fact that if perceptual demonstratives such as 'That tree' refer at all in the bad case, they do not have the same reference as in the good case.

5 Analogues of both sorts of sceptical scenario occur in Descartes's *Meditations*. He supposes that even in the most elementary reasoning, he may be confident that he is reasoning correctly when he is in fact reasoning fallaciously. Cartesian scepticism comprises scepticism about reason as well as scepticism about the external world.

vat is undergoing a massive illusion, why should that not include an illusion as to the extent of its evidence? After all, it is much worse placed for gathering evidence than it seems to itself to be. Although it still has the disposition to conform its beliefs to what the evidence supports, its unfortunate circumstances may interfere with its putting that disposition into effect. In particular, although it seems to itself to have exactly the same evidence as in the good case, it may in fact have less evidence. Its diminished stock of evidence may be insufficient for belief. For instance, if one's evidence is simply one's knowledge, and there is a knowledge norm for belief (believe *p* only if you know *p*!), then its violation of that knowledge norm amounts to a failure of rationality$_{\text{cont}}$, although not of rationality$_{\text{disp}}$. On that view, the brain has insufficient evidence to believe that it has hands. Of course, its failure is blameless, in a very specific way: it *appears* to itself to have sufficient evidence to believe that it has hands, but the appearance is deceptive.

Many epistemologists dismiss the idea that the brain in the vat suffers from a blameless failure of rationality. They typically argue that, in believing falsely that it has hands, the brain in the vat is not just blameless. It is following the very cognitive instincts it ought to have, doing what a well-designed brain should do. If instead it believed truly that it lacked hands, it would be following much worse cognitive instincts, since it has no evidence that it lacks hands. Even if it merely suspended belief, and became agnostic as to whether it had hands, it would be following a Pyrrhonian sceptical instinct that no well-designed brain should have, since in normal circumstances it involves self-imposing ignorance. In believing that it has hands, they argue, the brain is doing as well as it can. It is following exactly the same cognitive instincts as it follows in forming the same belief in the good case. They conclude that the belief is rational to exactly the same degree in the good and bad cases.

That line of argument implicitly focuses on rationality$_{\text{disp}}$. What it focuses on are the brain's dispositions. Such arguments have no force against the claim that the brain in the vat suffers from a blameless failure of rationality$_{\text{cont}}$.

In a recent paper, Stewart Cohen and Juan Comesaña treat it as obvious that rationality can require one to believe a falsehood. They give an example:

> Suppose you notice what appears to be a red table staring you in the face, you have no evidence of deception, everyone else around you says they see a red table, yet you fail to believe there is a red table before you. In our view you are paradigmatically irrational, even if unbeknown to you, you do not see a red table.[6]

Significantly, they apply the term 'irrational' to the agent. Clearly, if you are the sort of person who is generally disposed to think rationally and non-sceptically, and you exercise those dispositions in this case, then you will believe that there is a red table before you. Thus, in the circumstances, it is rational$_{\text{disp}}$ to believe (falsely) that there is a red table before you. But that does not mean that it is rational$_{\text{cont}}$ for you to have that belief.

Not all Cohen and Comesaña's claims about rationality fit rationality$_{\text{disp}}$. For instance, they say 'Rationality requires one to conform one's beliefs to one's evidence'.[7] That remark fits rationality$_{\text{cont}}$, not rationality$_{\text{disp}}$. Consider the variant of their example in which you do believe that there is a red table before you. That is just the sort of case in

6 Cohen and Comesaña (2013b: 407). In a footnote they add a qualification: 'One might hold that you are irrational only if you are taking some attitude toward whether there is a table.'

7 Cohen and Comesaña (2013a: 19).

https://doi.org/10.1017/epi.2017.24 Published online by Cambridge University Press

which someone with the disposition to conform their beliefs to what the evidence supports may very well misjudge what conforming their beliefs to their evidence currently involves. For since you are currently unaware of the perceptual illusion, you may easily but falsely take your current evidence to include facts about your environment of which you blamelessly but falsely take yourself to have perceptual knowledge. Thus it is just the sort of case in which rationality$_{disp}$ and rationality$_{cont}$ are in danger of coming apart.

For the sake of argument, Cohen and Comesaña explicitly leave externalist views of evidence open. In particular, of the equation E = K of one's total evidence with one's total knowledge, they say 'We are happy to grant E = K for the sake of argument' (2013b: 410). On such externalist views of evidence, one has less evidence in the bad case than in the corresponding good case. For instance, in the variant good case, you know that there is a red table before you, and your evidence includes the proposition that there is a red table before you. In the variant bad case, by contrast, you falsely believe that there is a red table before you, and your evidence does not include the proposition that there is a red table before you. Thus conforming your beliefs to your evidence in the bad case will involve something different from conforming your beliefs to your evidence in the good case. Consequently, rationality$_{cont}$ will also require something different in the bad case from what it requires in the good case, and it is rationality$_{cont}$ that is at issue when Cohen and Comesaña say that rationality requires one to conform one's beliefs to one's evidence. To leave externalist views of evidence genuinely open, one must take seriously the idea that what rationality (in the sense of rationality$_{cont}$) requires may differ sharply between the good and bad cases.

The argument in Section 2 against the equivalence of rationality$_{cont}$ and rationality$_{disp}$ did not depend on an externalist view of evidence; it works even on a phenomenalist conception. Contradictions have zero probability whatever the evidence. What the argument did depend on was *internalism* about dispositions. Switching the brain scrambler on or off was assumed to make no difference to Innocent's disposition to conform her beliefs to what her evidence supported, just as bubble-wrapping the vase makes no difference to its fragility. We do indeed tend to envisage dispositions as internal or intrinsic to the things so disposed. However, the tendency is not inexorable. Some dispositions are extrinsic (McKitrick 2003; Fara 2005). Castles became more *vulnerable* when gunpowder was introduced.

How do things look if we suppose that the brain scrambler causes Innocent to lose her disposition to conform her beliefs to what her evidence supports, even though she retains the intrinsic structures underlying that disposition? We then lose the argument that Innocent believes what someone disposed in those circumstances to believe what their evidence supports would believe in those circumstances. Consequently, we lose the argument that rationality$_{cont}$ and rationality$_{disp}$ come apart. However, by undermining the robustness of the dispositions underlying rationality, we also go some way towards undermining the assumption that the victims of more familiar sceptical scenarios retain their rationality. After all, the brain in a vat suffers from persistent massive hallucinations, a worrying sign for its state of mental health. If its cognitive dispositions changed in virtue of its envatment, can we still safely assume that it is rational$_{disp}$ to believe what it believes? If not, then given the equivalence schema it is also not safe to assume that it is rational$_{cont}$ to believe what the brain believes: in other words, it is not safe to assume that the brain's beliefs are supported by its evidence.

The issues are too large to settle here. I will rest content with two tentative conclusions. First, a potential source of internalism in epistemology (as exemplified by the claim that the subject's evidence is the same in the good and bad cases) is internalism about rationality, though only through the mediation of something like the equivalence schema. Second, once one rejects the equivalence schema and similar principles, and appreciates the distinction between $\text{rationality}_{\text{cont}}$ and $\text{rationality}_{\text{disp}}$, one may be able to do justice to normative assessments of rationality without assuming that false beliefs are ever adequately supported by the evidence.[8,9]

REFERENCES

Benton, M. 2013. 'Dubious Objections from Iterated Conjunctions.' *Philosophical Studies*, 162: 355–8.
Bird, A. 1998. 'Dispositions and Antidotes.' *Philosophical Quarterly*, 48: 227–34.
Cohen, S. and Comesaña, J. 2013a. 'Williamson on Gettier Cases and Epistemic Logic.' *Inquiry*, 56: 15–29.
—— **and** —— 2013b. 'Williamson on Gettier Cases in Epistemic Logic and the Knowledge Norm For Rational Belief: A Reply to a Reply to a Reply.' *Inquiry*, 56: 400–15.
DeRose, K. 2002. 'Assertion, Knowledge, and Context.' *Philosophical Review*, 111: 167–203.
—— 2009: *The Case for Contextualism*. Oxford: Clarendon Press.
Fara, M. 2005. 'Dispositions and Habituals.' *Noûs*, 39: 43–82.
Lasonen-Aarnio, M. 2010. 'Unreasonable Knowledge.' *Philosophical Perspectives*, 24: 1–21.
Lewis, D. 1997. 'Finkish Dispositions.' *Philosophical Quarterly*, 47: 143–58.
Martin, C. B. 1994. 'Dispositions and Conditionals.' *Philosophical Quarterly*, 44: 1–8.
McKitrick, J. 2003. 'A Case For Extrinsic Dispositions.' *Australasian Journal of Philosophy*, 81: 155–74.
Srinivasan, A. 2015. 'Are we Luminous?' *Philosophy and Phenomenological Research*, 90: 294–319.
—— ms. 'What's in a Norm?' Typescript.
Williamson, T. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.
—— 2013a. 'Gettier Cases in Epistemic Logic.' *Inquiry*, 56: 1–14.
—— 2013b. 'Response to Cohen, Comesaña, Goodman, Nagel, and Weatherson on Gettier Cases in Epistemic Logic.' *Inquiry*, 56: 77–96.

8 Cohen and Comesaña (2013b: 407) complain that in Williamson (2000), like them and in contrast to Williamson (2013b), I rely without argument on the view that one can rationally believe a falsehood. In the passages they quote from Williamson (2000), I am using the word 'rationally' in the sense of something like '$\text{rationally}_{\text{disp}}$'. By contrast, in Williamson (2013b) I use the term in the sense of something like '$\text{rationally}_{\text{cont}}$', in order to engage more directly with the statement in their previous paper 'Rationality requires one to conform one's beliefs to one's evidence' (Cohen and Comesaña 2013a: 19 in response to Williamson (2013a)). The terminological shift was confusing, and I should have warned the reader, but it did not signal any significant theoretical shift. Unlike 'knowledge', 'rationality' has never been one of my key theoretical terms, which is why I was so willing to adapt its use to the occasion in Williamson (2013b). In Williamson (2000), I argued for a view on which, *whatever* rationality is, we are sometimes rational without being in a position to know that we are being rational, and sometimes irrational without being in a position to know that we are being irrational, so that being in a position to know that one is being rational and not being in a position to know that one is being irrational are tempting alternative standards not equivalent to the original one, with a consequent destabilizing effect on the norm of rationality (see Srinivasan 2015 and ms for a defence of the underlying 'anti-luminosity' argument and further exploration of its normative consequences respectively).

9 Thanks to Stewart Cohen and David Sosa for extensive written comments on an earlier version of this paper, and to participants in the 2016 Episteme Conference in Skukuza for helpful discussion.

https://doi.org/10.1017/epi.2017.24 Published online by Cambridge University Press

—— 2016. 'Knowing by Imagining.' In A. Kind and P. Kung (eds), *Knowledge through Imagination*, pp. 113–23. Oxford: Oxford University Press.
—— Forthcoming. 'Justifications, Excuses, and Sceptical Scenarios.' In J. Dutant and F. Dorsch (eds), *The New Evil Demon*. Oxford: Oxford University Press.

TIMOTHY WILLIAMSON has been the Wykeham Professor of Logic at Oxford University since 2000. Before that he was a Lecturer in Philosophy at Trinity College Dublin (1980–88), Fellow of University College Oxford (1988–94), and Professor of Logic and Metaphysics at Edinburgh University (1995–2000). His publications include *Identity and Discrimination*, *Vagueness*, *Knowledge and its Limits*, *The Philosophy of Philosophy*, *Modal Logic as Metaphysics*, *Tetralogue*, and about 200 academic articles on logic, metaphysics, epistemology, philosophy of language, and metaphilosophy. He has held visiting positions at MIT, ANU, Canterbury University (NZ), Princeton, UNAM (Mexico), Chinese University of Hong Kong, University of Michigan, and Yale. He is an Honorary Member of the Royal Irish Academy, Fellow of the British Academy, Fellow of the Royal Society of Edinburgh, Foreign Member of the Norwegian Academy of Science and Letters, Foreign Honorary Member of the American Academy of Arts and Sciences, and Member of the Academia Europaea.