

If You Can't Change What You Believe, You Don't Believe It

GRACE HELTON
Princeton University

Abstract

I develop and defend the view that subjects are necessarily psychologically able to revise their beliefs in response to relevant counter-evidence. Specifically, subjects can revise their beliefs in response to relevant counter-evidence, given their current psychological mechanisms and skills. If a subject lacks this ability, then the mental state in question is not a belief, though it may be some other kind of cognitive attitude, such as a supposition, an entertained thought, or a pretense. The result is a moderately revisionary view of belief: while most mental states we thought were beliefs are beliefs, some mental states which we thought were beliefs are not beliefs. The argument for this view draws on two key claims: First, subjects are rationally obligated to revise their beliefs in response to relevant counter-evidence. Second, if some subject is rationally obligated to revise one of her mental states, then that subject can revise that mental state, given her current psychological mechanisms and skills. Along the way to defending these claims, I argue that rational obligations can govern activities which reflect on one's rational character, whether or not those activities are under one's voluntary control. I also show how the relevant version of epistemic 'ought' implies 'can' survives an objection which plagues other variants of the principle.

Key words: nature of belief; doxastic obligation; norms of rationality; 'ought' implies 'can'; doxastic voluntarism; rational character; epistemic normativity; delusion; implicit bias

If you believe there's coffee made but then, walking into the kitchen, see that the coffee pot is empty, you will stop believing there's coffee made and will start believing there's no coffee made. If you believed the U.K. would not vote to leave the European Union but then read the news the day after the vote, you stopped believing the U.K. would not vote to leave the E.U. and started believing the U.K. would—and indeed, did—vote to exit the E.U. If you believe Missoula is the capital of Montana but then your native Montanan friend tells you that Helena is her home state's capital, you will stop believing Missoula is Montana's capital and will start believing Helena is that state's capital.

These prosaic cases suggest the following straightforward connection between belief and evidence: Subjects form and revise their beliefs in accordance with the

evidence. Or, if subjects do not invariably regulate their beliefs in response to the evidence, they are strongly so inclined. The descriptive claim that subjects are strongly inclined to regulate their beliefs in accordance with the evidence is different than the normative claim that subjects *ought* to regulate their beliefs in this way. The descriptive claim says that if a subject is not strongly inclined to regulate a certain mental state in this way, then that mental state is not a belief, though it may be some other kind of cognitive attitude, such as an entertained thought, a pretense, or a non-doxastic delusion. The descriptive claim is widely accepted among theorists of belief, though it is not without detractors.¹

The view that subjects are strongly inclined to regulate their beliefs in accordance with the evidence stands in tension with certain empirical results. For instance, the view stands in tension with empirical evidence that subjects by default believe whatever propositions they entertain, regardless of whether those propositions are well-supported (Mandelbaum 2014). The view also stands in tension with the result that subjects are inclined to retain their existing beliefs, even when those subjects have counter-evidence for those beliefs (Nickerson 1998). The worry looms that if subjects are necessarily strongly inclined to regulate their beliefs in accordance with the evidence, then very few mental states will count as beliefs.

In this paper, I argue for a different connection between belief and evidence, one which obtains whether or not subjects are necessarily strongly inclined to regulate their beliefs in response to the evidence. The connection is this: subjects necessarily have a certain robust psychological *ability* to revise their beliefs in response to relevant counter-evidence. Specifically, subjects are able, given their current psychological mechanisms and skills, to revise their beliefs in accordance with the evidence. If a subject cannot revise a mental state in this way, then the mental state in question is not a belief, though it may be some other kind of cognitive attitude, such as a supposition, an entertained thought, or a pretense. I dub the view that beliefs are subject to this constraint *the revisability view of belief* (or *the revisability view*, for short). Since ability and inclination can go together but can also come apart, the revisability view is consistent with but does not entail the view that subjects are necessarily strongly inclined to regulate their beliefs in response to the evidence.

Note that the revisability view concerns only how subjects revise existing beliefs, not how subjects *form* beliefs. Thus, the view is compatible with empirical evidence that subjects form beliefs by default believing propositions they entertain, whether or not those propositions are well-supported (Mandelbaum 2014). The view is also compatible with the result that subjects are sometimes at most weakly inclined to revise their beliefs in response to relevant counter-evidence. The view requires only that subjects be able to revise their beliefs, not that they be strongly inclined to do so.

The fact that the revisability view can accommodate irrational beliefs shows merely that the view crosses a hurdle which any view of belief must ascend. It does not suggest a positive reason to accept the view. The centerpiece of the paper is such an argument, which I dub *the argument from the norm of revision*. This argument moves from the claim that subjects are rationally obligated to revise their

beliefs in response to relevant counter-evidence to the conclusion that subjects have a psychological ability to revise their beliefs in response to such evidence. The key to this transition is a certain epistemic version of the principle 'ought' implies 'can.' Painted in broad strokes, the argument is as follows:

- (1) Necessarily, if some subject believes p and has sufficiently strong, undefeated evidence that $not-p$, then that subject has a *prima facie* obligation of rationality to revise that belief.
- (2) Necessarily, if some subject has a *prima facie* obligation to revise one of her mental states, then that subject has a certain robust psychological ability to revise that mental state.
- (3) Necessarily, if some subject believes p and has sufficiently strong, undefeated evidence that $not-p$, then that subject has a certain robust psychological ability to revise that belief.

Two points about the argument are worth highlighting at the outset: First, the argument's conclusion alone tells us nothing about how ubiquitous beliefs are within human psychology. If it should turn out that humans can revise very few or even none of their mental states in the way relevant way, the conclusion would entail that humans have very few or no beliefs.

I take the desideratum that humans have many beliefs to be fundamental to our theorizing about belief. That is, I think we should reject any view which entails that humans have few or no beliefs. As it turns out, the revisability view does not threaten this desideratum. For, as I have already suggested, there is empirical evidence that humans *can* revise a wide range of their mental states in the relevant way, even if humans seldom do revise their beliefs in this way (I say more about this in §2.1).

Second, notice that the argument's conclusion is consistent with the possibility that subjects might have beliefs for which there is no actual or even possible counter-evidence. Subjects count as trivially capable of revising such states, if any, in virtue of never failing to revise those states in response to relevant counter-evidence. So, the view permits such mental states into the class of belief (I elaborate on this aspect of the view in §1).

Here is the plan for the paper: I first more fully characterize the revisability view and draw out some of its consequences (§1). I next develop and defend the main argument for the revisability view, paying special attention to the worry that if subjects cannot voluntarily regulate their beliefs, then we should reject: the norm of revision, 'ought' implies 'can', or the conjunction of these claims (§2). I then defend the revisability view against the following objection: the view excludes all unrevisable mental states from the class of belief, even when subjects sincerely endorse those mental states, rely on them in action, and exploit them in a wide range of inferences. Such mental states are intuitively beliefs, and therefore, the revisability view is too strong. I argue that, on balance, we should welcome the result that such mental states are not beliefs (§3). I close with a brief summary (§4).

Before proceeding to the first section, a methodological point may be useful. The proposed account of belief is not intended as a theory of how reasoners in fact use the concept of belief. Rather, the view attempts to capture what is at the core of the concept of belief, i.e., which platitudes about belief are indispensable. I find that the platitude that beliefs are governed by certain norms of rationality survives scrutiny. I find that certain other platitudes about belief—such as the platitude that mental states which are inferentially promiscuous are necessarily beliefs—wither under scrutiny. The outcome is a moderately revisionary approach to belief: while many states we thought were beliefs are beliefs, at least some mental states we thought were beliefs are not beliefs.

Finally, it may be worth saying something about the import of the paper's thesis for issues outside of epistemology. Most saliently, the present thesis has the potential to reconfigure a spate of recent disputes concerning the nature of certain puzzling mental states, such as clinical delusions, disavowed prejudices, and religious convictions.² In debates about the nature of these mental states, the question invariably arises whether the state in question is a belief or something else, such as a pretense. Often, those who argue that such states are not beliefs point to the (putative) fact that subjects are not strongly inclined to regulate such states in response to the evidence, whereas—it is claimed—subjects are inclined to regulate their beliefs in this way.

If, as I have suggested, there are reasons to doubt the claim that subjects are strongly inclined to regulate their beliefs in response to the evidence, then there are reasons to doubt whether subjects' inclination to regulate (say) their disavowed prejudices is relevant to the question of whether those prejudices are beliefs. At the same time, the revisability view suggests a different connection between belief and evidence, one which might help recast such disputes. Instead of asking whether subjects are inclined to regulate (say) their disavowed prejudices in accordance with the evidence, we should ask whether subjects have the relevant psychological ability to revise their prejudices in accordance with the evidence. If subjects lack this ability, then their prejudices are not beliefs. If subjects have this ability, then it remains a possibility that such prejudices are beliefs. The same points apply to clinical delusions and religious convictions, respectively.

1 The Revisability View of Belief

In this section, I elaborate on the revisability view and draw out some of its consequences. Here is the view in full:

THE REVISABILITY VIEW OF BELIEF

Necessarily, if some subject's mental state that *p* is a belief and if that subject has sufficiently strong, undefeated evidence that *not-p*, then that subject is able to revise that mental state, given her current psychological mechanisms and skills.

On this view, subjects can revise their beliefs, given the psychological mechanisms and skills they already possess. So, it is not enough for some subject's mental state to count as a belief that that subject could revise that mental state, were she to

acquire new psychological mechanisms or skills. This subject's current repertoire of psychological mechanisms must enable her to revise that state.

For some subject to count as revising one of her mental states in response to counter-evidence, she must reduce the strength of that mental state, suspend that mental state, or eliminate that mental state altogether.³ While this revision must be something the subject does, it need not be under her voluntary control. Rather, belief revision might be like many other activities which we do but which aren't under our voluntary control. These include recognizing our loved ones by their voices, falling asleep, and visually distinguishing between green and blue.⁴

Importantly, the claim that a given subject can revise any of her beliefs does not entail that that subject can *simultaneously* revise all of her beliefs. This restriction is consonant with how we think of other abilities: the fact that Janelle can swim a mile and can also play the *Rhapsody in Blue* clarinet solo does not entail that Janelle can swim a mile *while* playing the *Rhapsody* solo.

It is worth stressing that the revisability view is not a full account of belief. Rather, the view posits only a necessary condition on belief, which means that the view is consistent with several other proposed necessary conditions on belief. For instance, the revisability view is consistent with the claim that subjects are necessarily disposed to act in accordance with their beliefs (Schwitzgebel 2001, 2002, 2010). The view is also consistent with the claim that beliefs are necessarily *inferentially promiscuous*, or available as a premise across a wide range of inferences (Glüer & Wikforss 2013b, Mandelbaum 2014, Stich 1978, 507). At the same time, the view is at odds with the suggestion that behavioral dispositions or inferential promiscuity are sufficient for belief. (I defend this implication of the view in §3).

One natural question about the revisability view is what the view says about cases in which a subject lacks counter-evidence for what she believes. There are two ways this might happen, and the view treats them differently. First, some subject might, for merely contingent reasons, lack counter-evidence for one of her views. For instance, the subject who incorrectly represents that Missoula is the capital of Montana might, merely by happenstance, never encounter evidence to the contrary. On the revisability view, this representation can count as a belief so long as this subject *would* revise her view, *were* she to encounter relevant counter-evidence.

Second, some subject might, of necessity, lack counter-evidence for one of her views. For instance, suppose a subject has the view that nothing is both round and square. It is at least plausible that there is no actual or possible evidence which might tell against this view. The revisability view permits all such mental states (if any) into the class of belief, because subjects who hold these views never fail to revise them in response to relevant counter-evidence. That is, subjects are trivially capable of revising such states, so these states can qualify as beliefs.⁵

I turn next to the argument for the revisability view.

2 The Argument from The Norm of Revision

The argument I offer in favor of the revisability view is *the argument from the norm of revision*. This argument is meant to extend in full generality to all doxastic states,

whether occurrent or dispositional, attended or unattended, unconsidered or reflectively endorsed, conscious or non-conscious, compartmentalized from other states or integrated with other states, heuristically-produced or inferentially-produced. Thus, the argument is meant to extend both to beliefs and to what are sometimes called judgments.

The argument is named after its first premise, the claim that subjects are rationally obligated to revise their beliefs in accordance with counter-evidence. Since this claim is normative, it cannot by itself illuminate the descriptive nature of belief. But combining this claim with a certain epistemic variant of the principle ‘ought’ implies ‘can’ yields an argument in favor of the revisability view. Here is the argument in full:

THE ARGUMENT FROM THE NORM OF REVISION

- (1) Necessarily, if some subject believes *p* and has sufficiently strong, undefeated evidence that *not-p*, then that subject has a *prima facie* obligation of rationality to revise that belief in accordance with the evidence.
- (2) Necessarily, if some subject has a *prima facie* obligation of rationality to revise some mental state, then that subject can, given those psychological mechanisms and skills which she already possesses, revise that mental state in the relevant way.
- (3) Necessarily, if some subject believes *p* and has sufficiently strong, undefeated evidence that *not-p*, then that subject can, given those psychological mechanisms and skills which she already possesses, revise that belief in accordance with the evidence.

Since the argument extends to arbitrarily chosen subjects and to arbitrarily chosen beliefs, its conclusion is that for any given subject and for any of that subject’s beliefs, that subject has a certain robust psychological ability to revise that belief in response to the evidence. Specifically, that subject can revise her belief, given her current psychological mechanisms and skills. If a subject lacks this ability, then the mental state in question is not a belief, though it may be some other cognitive attitude, such as an assumption, a merely entertained thought, or a pretense.

2.1 Does This Style of Argument Prove Too Much?

Both of the argument’s premises are controversial, and I will defend each momentarily. Before doing so, it may be worth defusing a different objection, one directed at the overall style of argument. According to this objection, there must be *something* wrong with the argument because there are structurally similar arguments with implausible consequences. Consider, for instance, the following argument, which is similar to the argument from the norm of revision except that it is centered around a norm regulating the formation of belief, instead of the revision of belief:

THE (UNSUCCESSFUL) ARGUMENT FROM THE NORM OF FORMATION

- (4) Necessarily, subjects ought to form their beliefs in response to good evidence.
- (5) Necessarily, if some subject ought to form some mental state in response to good evidence, then that subject can form that mental state in response to good evidence.
- (6) Necessarily, subjects can form their beliefs in response to good evidence.

The problem with this argument is that its conclusion can be combined with a certain empirically plausible view about belief formation to yield the result that humans have few or even no beliefs. The empirically plausible view to which I refer is the view—already mentioned in the paper's introduction—that humans by default believe all or nearly all of the propositions they entertain, whether or not those propositions are supported by the evidence (Mandelbaum 2014). I am not certain whether this view is correct, but for the sake of argument, let's suppose that it is. Let's further suppose that this view has the implication that humans are not *capable* of forming beliefs in accordance with good evidence. Combining this view with the argument's conclusion yields the implausible result that humans have no beliefs. For, if humans cannot form any of their mental states in accordance with the evidence and yet, subjects must form beliefs in this way, no human mental states will count as beliefs.

I think we should reject any view which has the consequence that humans have few or no beliefs. Thus, I accept that there is something wrong with the argument from the norm of formation. On my diagnosis, the argument is valid but unsound, because its first premise is false. The norm of formation turns out to be false precisely because it would force us to the view that humans have no or very few beliefs. Put otherwise, the so-called norm of formation is not really a norm. It is a pseudo-norm, i.e., a proposed norm which has all the initial appeal of a genuine norm but whose implications are sufficiently unacceptable that it is no norm at all.⁶

Unlike the norm of formation, the norm of revision does not have the consequence that humans have few or no beliefs. For, even if humans cannot *form* their views in accordance with the evidence, there is reason to think that humans can *revise* a very wide range of their views in accordance with the evidence. For instance, it has been shown that subjects moderate their views when they are encouraged to consider evidence which runs counter to those views (Schwind et al. 2012). In addition, the mechanism of *cognitive dissonance* helps to regulate a very wide range of human mental states, arguably in an evidence-responsive way (Harmon-Jones 2000, Harmon-Jones & Harmon-Jones 2007). Finally, there is some evidence that subjects can revise heuristically-produced judgments, or what are sometimes termed System 1 outputs, even if such outputs are not produced in a rational way and even if subjects do not tend to revise such judgments (Svedholm-Häkkinen 2015, Thompson & Johnson 2014).⁷

So, while there are defective arguments that are structurally similar to the argument from the norm of revision, this fact does not suggest that there is something defective about the argument from the norm of revision. For, if the argument from

the norm of formation is any guide, the reason these structurally similar arguments fail is that they are valid but unsound, in virtue of their reliance on pseudo-norms. In contrast, the premises in the argument from the norm of revision are both true, as I will now show.

2.2 *The Norm of Revision*

The first premise of the argument just is the norm of revision. It says that the subject who believes p and who has sufficiently strong, undefeated evidence that $\text{not-}p$ has a *prima facie* obligation to revise her belief in accordance with that evidence. The question of how strong evidence must be to be sufficiently strong will depend on the ultimate correct theory of how subjects ought to update their beliefs in response to evidence.

The obligation in question is *prima facie* in the sense that it can be trumped by other obligations—whether obligations of morality, prudence, or other obligations of rationality. When *prima facie* obligations are trumped, they do not cease to exist. Rather, such obligations retain their force even when they are not, all things considered, what their subject ought to do.⁸

In favor of the norm of revision is that it explains several ordinary but powerful intuitions. For instance, if you believe there is fruit on your kitchen table and then, walking into the kitchen, see that the fruit bowl is empty, you should revise your belief. If you believe your child did not shoplift from a local convenience store and then view surveillance footage showing her doing just that, you should revise your belief. If you believe that Norway is a member of the E.U. and then are informed by your Norwegian friend that his home country is not part of the E.U., you should revise your belief.

The claim that in such cases you have a rational obligation to revise your beliefs is consistent with the claim that morality or prudence might recommend different courses of action. For instance, it might be that morality requires that you believe your child when she says that she did not shoplift, even if the surveillance footage says otherwise. The presence of such an overriding moral obligation would not make it the case that you no longer have a *prima facie* obligation of rationality to revise your belief. In such a case, the moral obligation merely trumps your obligation of rationality such that this obligation is not, all things considered, what you ought to do.

Importantly, the norm of revision extends in full generality to all beliefs, including non-conscious beliefs and beliefs formed on the basis of perception. Suppose you judge, on the basis of a visual experience, that two lines in a figure are of different lengths. If you should come to possess good evidence that this figure is illusory in just this respect, you should revise your belief. Or, suppose you come to believe implicitly, as the undetected effect of watching too many commercials, that drinking fruit juice is healthy. If you should come to possess good evidence that fruit juice has deleterious effects on health, you should revise your belief.

Despite the intuitive appeal of the norm of revision, there are several reasons to reject it. One reason comes from the view that if beliefs are not under voluntary control, then there are no obligations of belief. Another reason comes from the

view that norms of rationality, if any, must be action-guiding, but that no putative norms of rationality are action-guiding. I will consider and criticize each of these reasons for rejecting the norm of revision.

2.2.1 An Objection from The View That Beliefs Are Not Under Voluntary Control

The first objection to the norm of revision runs as follows: Activities governed by obligations are necessarily under voluntary control. Beliefs are not under voluntary control. So, beliefs are not governed by obligations. Hence, the norm of revision, which posits such an obligation, is false (Alston 1988).

My strategy here is to remain neutral on the question of whether beliefs are under voluntary control.⁹ Instead, I will criticize the claim that activities governed by obligations are necessarily under voluntary control. I will consider and reject two motivations for this claim.

2.2.1.1 The View That Blameworthiness Entails Voluntary Control

One reason to think that activities governed by obligations must be under voluntary control derives from a plausible view of the relation between obligation and blameworthiness. On this view, if some subject fails to carry out an all things considered obligation, then she is necessarily blameworthy for that failure. If this is right and if activities which are obligated are sometimes not under voluntary control, then there will be cases in which subjects are blameworthy for activities not under their voluntary control. This result has seemed unacceptable to many (Rusin 2010).

I will argue that we should reject the initially appealing view that subjects are blameworthy only for those activities which are under their voluntary control. In its place, I defend the view that subjects are blameworthy only for those activities which reflect on their *character*, i.e., on their desires, values, or cares. On this view, there can be obligations of belief even if beliefs are not under voluntary control. The reason one can be rationally blameworthy for failing in these obligations is that these failures reflect on one's *rational character*, i.e., on whether and how much one desires, values, or cares about epistemic goods, such as truth.¹⁰

The view that subjects can be blameworthy only for those activities under their voluntary control is motivated by intuitive contrasts between cases. If you kick someone out of spite, you are blameworthy for doing so. If you kick someone merely because you are having a seizure, you are not blameworthy for doing so. One natural explanation of this difference is that the spiteful kick is under your control, whereas the seizure-driven kick is not. This suggests that one can be blameworthy for some activity only if that activity is under one's voluntary control. Relatedly, one can be obligated to perform some activity only if that activity is under one's voluntary control.

Though the voluntarist explanation of blameworthiness can account for the difference in the 'kicking' cases, the character-based view can also explain this contrast. On this view, what explains the difference in these cases is that the spiteful kick reflects poorly on one's character—i.e., on one's desires, values, or cares—whereas the seizure-driven kick does not. The spiteful kick suggests that one does

not care enough about others' well-being, whereas the seizure-driven kick suggests nothing whatsoever about what one cares about.

Since both the voluntarist and character-based views of blameworthiness can explain the contrast between cases like the 'kicking' cases, we need some further way of adjudicating between the views. Toward this end, I will draw on the work of several theorists to argue that: there are obligations of emotion, we can be blameworthy for flouting these obligations, and these obligations exist whether or not emotions are under our voluntary control (Adams 1985; Chuard & Southwood 2009, 614–625; McHugh 2012, 87–89).¹¹ I will then show that while the voluntarist view of blameworthiness cannot accommodate obligations of emotion, the character-based view can.

Intuitively, you should feel outrage in response to deep injustice, gratitude for undeserved kindness, and remorse when you have needlessly hurt others (Adams 1985; Card 1988; Chuard & Southwood 2009, 614–625; Hieronymi 2008, 362; McHugh 2012, 87–89). These obligations seem to concern emotions themselves, not attendant actions. Consider, for instance, the person whose friend feels no remorse for hurting her. This person might be justifiably more injured by her friend's callousness than by her friend's initial wrongdoing. Moreover, this person might justifiably blame her friend more for her lack of remorse than for the initial wrongdoing. Thus, her friend's lack of remorse seems *itself* to be blameworthy and to *itself* go against what friendship requires (McHugh 2012, 87; see also Adams 1985, 4–6). If we take these intuitions at face value, then there are obligations to feel certain emotions in certain circumstances.

One way to resist the claim that there are obligations of emotion is to claim that there are merely obligations to act as though one has certain emotions. For instance, when one hurts a friend, one should act as though one is remorseful, but one need not feel remorse (Adams 1985, 4). To show that this suggestion cannot be right, consider again the wronged friend. Suppose this person were to discover that the friend who had wronged her had then proceeded to *feign* remorse. This person might be justifiably hurt by her friend's failure to feel remorse, despite her friend's efforts to act remorseful. This suggests that merely acting as though one feels remorse does not satisfy the obligation in question. One must in fact feel remorseful.

There is reason, then, to think that there are obligations of emotion, obligations which, when flouted, can result in one's being blameworthy. At the same time, it is at least not obvious that emotions are under voluntary control. Suppose someone offers you a large financial reward to feel gratitude for a kindness shown to you by a person you have long loathed. Could you collect the money? It seems at least doubtful whether you could. In contrast, suppose someone offers you a large financial reward to raise your right hand. Could you collect the money? It seems very likely that you could.¹² The difference between these cases suggests that it is at least not clear whether emotions are under voluntary control.

It might be suggested that even if emotions are not under the same sort of control which governs bodily actions, emotions can be controlled by other, indirect means. For instance, perhaps by focusing your attention on the benefits conferred

on you by your enemy's kindness, you might coax yourself into feeling gratitude to her. However, even if this sort of indirect strategy might succeed in some cases, there are likely other cases in which it is doomed to fail. Extenuating circumstances and the psychology of the subject in question likely play a role in whether indirect manipulations will have any chance of success.

Here, then, we have two facts: First, there is the fact that there are obligations of emotion. Second, there is the fact that it is *not clear* whether emotions are under voluntary control. I take the combination of these facts to support the following claim: our intuition that there are obligations of emotion is not hostage to the view that emotions are under voluntary control. For if it were, given the lack of clarity about whether emotions are under voluntary control, we would not have this intuition.

My claim is the following counterfactual one: were it to turn out that emotions are not under any kind of voluntary control, and were this fact widely known, this would not diminish the conviction that subjects should feel outrage at deep injustice, gratitude for undeserved kindness, and remorse for needlessly hurting others. I take this (counterfactual) fact to be best explained by another (counterfactual) fact: there would still be obligations of emotion, even if emotions were not under voluntary control. This counterfactual claim is enough to falsify the view that one can have obligations concerning only those activities which are under one's voluntary control.¹³ On the presumption that failing to satisfy an obligation renders one blameworthy, this result entails that one can be blameworthy for activities not under one's voluntary control.

It is in accounting for obligations of emotion that the character-based view of blameworthiness outperforms the voluntarist view. For the character-based view can explain why emotions can be governed by obligations—obligations which, when flouted, result in blameworthiness—even if emotions are not under voluntary control. For instance, feeling remorse for hurting others reflects that one cares about others' well-being. Feeling outrage in response to a minor slight betrays that one cares too much about one's own respect. In this way, emotions can reflect on the moral character of a person, even if those emotions are not under that person's voluntary control.¹⁴

Once we have the character-based view of blameworthiness in hand, we are in a position to explain why one can have rational obligations concerning one's beliefs, even if one's beliefs are not under one's voluntary control. This is because beliefs can reflect on one's rational character, i.e., on whether and how much one values, desires, or cares about epistemic goods, such as truth and knowledge. For instance, it reflects positively on one's rational character if one believes on good evidence that whether some coin will come up heads does not depend on whether that coin came up heads in the previous toss. It reflects negatively on one's rational character if one believes on poor evidence that astrology is a good predictor of future outcomes. One can be praiseworthy for the first belief insofar as it reflects that one cares about truth or other epistemic goods. One can be blameworthy for the second belief insofar as it reveals that one cares too little about truth or other epistemic goods.¹⁵

I conclude that considerations from blameworthiness do not support the view that there are no obligations of belief.

2.2.1.2 *The View That ‘Ought’ Implies ‘Voluntarily Can’*

I now turn to a different argument for the view that if beliefs are not under voluntary control, then there are no obligations of belief. This argument runs as follows: first, ‘ought’ implies ‘can,’ i.e., if some subject ought to perform some activity, then that subject can perform that activity. Second, ‘can’ implies ‘voluntarily can,’ i.e., if some subject ought to perform some activity, that then subject can voluntarily perform that activity. So, ‘ought’ implies ‘voluntarily can.’ If this is right, and if beliefs are not under voluntary control, then there are no obligations of belief. Hence, the norm of revision, which posits such an obligation, is false.

I reject the argument’s second premise, the claim that ‘can’ entails ‘voluntarily can.’ As both Conor McHugh and the team of Philippe Chuard and Nicholas Southwood have argued, it is a mistake to think that if some subject can perform some activity, then that subject can *voluntarily* perform that activity (Chuard & Southwood 2009, 614–625; McHugh 2012, 86–93). Subjects can do all sorts of things not under their voluntary control. They can sneeze and can hiccup. They can fall asleep and can dream. They can hear sounds out their window and can see objects in their line of sight. They can tell whether an animal before them is a dog or a cat. They can understand spoken language. Though all of these are things subjects *can* do, none of these are things subjects *voluntarily* do.¹⁶ So, ‘ought’ implies ‘can’ does not support the view that if beliefs are not under voluntary control, then there are no obligations of belief.

2.2.2. *An Objection from the View that Obligations Must Guide Action*

Recent work by Kathrin Glüer and Åsa Wikforss suggests a different argument against the norm of revision and against norms of rationality more broadly. This argument runs as follows: norms are necessarily action-guiding. So, if there are norms of rationality, then subjects must sometimes revise their beliefs in a way that is guided by those norms and not merely in conformity with them. But, it is implausible that subjects are ever guided by such norms, even when their beliefs do conform to them. So, it is implausible that there are norms of rationality (Glüer & Wikforss 2013a).¹⁷

I reject the argument’s first premise, the claim that norms of rationality, if any, must guide action. On the face of it, this might seem like a strange move. For what would be the point of positing such norms if they do not—perhaps, *could* not—guide action? This is a large and fraught question about the nature of normativity and not one I could hope to settle here. But to briefly sketch my own view: even though norms of rationality are not action-guiding, they earn their keep in others ways. For instance, they serve as a useful theoretical tool for making sense of rational blameworthiness. What it is to say that someone is rationally blameworthy is to say that she fails to conform to some all things considered norm of rationality. Such an explanation can both clarify the nature of rational blameworthiness and help us identify novel instances of its kind. On this view, a norm is a measuring

tape, not a whip. Thus, even if it should turn out that the (putative) norm of revision is not action-guiding, this would not be a reason to deny that this norm exists.

In this section, I have considered two objections to the norm of revision and have argued that neither succeeds. In light of this result, and in light of the intuitions which motivate the norm of revision, I conclude that we should accept this norm.

I turn to defending the argument's second premise.

2.3 Epistemic 'Ought' Implies A Robust Psychological 'Can'

In this section, I make two separable claims: First, I defend the argument's second premise, which is a variant of 'ought' implies 'can,' one on which one's rational obligations are constrained by one's current psychological mechanisms and skills. Second, I argue against a different version of 'ought' implies 'can,' one on which one's rational obligations are constrained by one's current mental states. The latter result will be useful in defending the preferred variant of 'ought' implies 'can' from a certain objection.

Here again is the argument's second premise:

EPISTEMIC 'OUGHT' IMPLIES A ROBUST PSYCHOLOGICAL 'CAN'

Necessarily, if some subject has a *prima facie* obligation of rationality to revise some mental state, then that subject can, given those psychological mechanisms and skills which she already possesses, revise that mental state in the relevant way.

As argued in the previous section, a subject might be able to revise a mental state even if she cannot voluntarily revise that mental state. Likewise, a subject might be able to fall asleep or recognize her loved ones on sight, even if she cannot do those things voluntarily.

Psychological mechanisms and skills include both idiosyncratic and more widely held capabilities. They also span the divide between innate and learned traits. Examples include: the ability to empathize with others, the ability to read Mandarin, the ability to do calculus, the ability to discriminate between heard pitches, the ability to see movement, and the ability to resolve dissonant attitudes in the direction of greater mutual coherence.

For our purposes, it will suffice to analyze psychological mechanisms and skills in terms of counterfactual behavior. On this view, in order for a subject to count as possessing some psychological mechanism or skill, she must exhibit whatever trait or activity exemplifies that mechanism or skill in some range of those counterfactual circumstances in which her overall psychology is similar to its current state. For instance, in order for a subject to count as able to read Mandarin, she must correctly interpret Mandarin in some range of those counterfactual circumstances in which her overall psychology is similar to its current state. It is not enough for this subject to count as able to read Mandarin that she might, by some freak coincidence, offer a correct translation of a Mandarin text. This 'fluke-ish' behavior is not exhibited in a sufficiently wide range of counterfactual circumstances to count as an ability. Likewise, it is not enough for some subject to count as able to read Mandarin that

she might take some courses and thereby come to read the language. Counterfactual circumstances in which this subject has undergone substantial training in Mandarin are also ones in which that subject's psychology is quite different from its present state and so, such circumstances are irrelevant to the question of whether this subject is currently able to read the language.¹⁸

Psychological mechanisms and skills form a natural contrast with *mental states*, such as thoughts, moods, desires, sensations, and perceptual experiences. (This difference will be important shortly). The belief that it is raining out is neither a mechanism nor a skill, though it may be produced by or governed by mechanisms or skills. The same can be said of a feeling of exuberance, a searing pain in one's leg, a desire for pistachio ice cream, or a distracted mood. This ostensive characterization of mechanisms and skills, on the one hand, and of mental states, on the other, is sufficient for our purposes.

The relevant 'ought' implies 'can' principle is motivated by intuitions about certain cases. For instance, someone who is fluent in English but who has no grasp of Mandarin cannot be rationally required to revise her inaccurate belief about what some text means just by looking at its Mandarin translation. In certain circumstances, someone who reads Mandarin fluently might be so required. Likewise, someone who lacks the ability to identify pitches by their sound cannot be rationally required to revise her belief about which key a certain song is in. Someone who has perfect pitch might be so required.

Reflection on these cases suggests that rational obligation is constrained by one's current abilities, not (merely) by those abilities one might acquire. For, if rational obligation were constrained merely by those obligations one might acquire, the person who doesn't read Mandarin might be rationally required to revise her belief about a Mandarin-encoded text. For, this person might learn Mandarin and so, Mandarin is among those abilities she might acquire.

The question arises: why aren't non-readers of Mandarin rationally required to revise their assessments on the basis of Mandarin-encoded information, whereas their Mandarin-fluent peers might be so required? And why aren't those who lack perfect pitch rationally required to revise their beliefs about which key a song is in, when those of their peers who are blessed with perfect pitch might be so required? If we accept (2) and accept that rational obligations are constrained by current psychological ability, we enjoy a straightforward explanation of these asymmetries: They are explained by differences in the relevant subjects' current psychological abilities.

It might be tempting to offer a different explanation of the contrasts in the 'Mandarin' and 'perfect pitch' cases. On this explanation, the reason (say) a Mandarin reader might be required to revise her beliefs in response to Mandarin-encoded information, whereas her non-Mandarin counterpart will not be so required, is due to differences in evidence, not due to differences in abilities *per se*. For the Mandarin reader is in possession of evidence which her non-Mandarin-reading counterpart lacks, and perhaps this difference alone explains their different rational requirements.

While it is true that the contrast in (say) the 'Mandarin' cases might be explained by differences in evidence, consideration of a different case breaks the tie between the evidence-based explanation and the 'ought'-implies-'can'-based explanation of such contrasts. Suppose a subject is in the middle of a long mathematical proof. She has, in some sense, all of the evidence she needs to reach some very difficult conclusion. However, she is working at the absolute limit of her cognitive processing capacity and cannot, psychologically speaking, draw the relevant conclusion. It seems to me that this subject is not rationally obligated to draw this conclusion, even though she is in possession of the relevant evidence. The claim that rational obligation is constrained by psychological ability can explain this result, whereas evidence-based considerations cannot.¹⁹ Thus, the view that rational obligation is constrained by psychological ability is the only explanation which can account for all of the cases considered.

So far, I've argued that a subject's rational obligations are constrained by that subject's current repertoire of psychological mechanisms and skills. The question arises whether a subject's rational obligations are also constrained by that subject's current mental states. That is, if you have some mental state which renders you unable to revise some belief, might you nevertheless be rationally obligated to revise that belief? I find that you might. Suppose that you incorrectly believe that $8 \times 6 = 46$. While splitting a bill, you rely on this belief and, as a result, come to an inaccurate view about how the bill should be split. The question is, rationally speaking, are you required to revise your belief about how the bill should be split? Let's suppose that, holding fixed your belief that $8 \times 6 = 46$, you cannot revise your belief about the bill. It seems to me that you are rationally required to revise your belief, your troublesome belief notwithstanding.

Or, to take another case, suppose that you are a juror in a criminal trial, one in which the evidence suggests that the defendant is innocent of the crime of which he is accused. You find the defendant to be a repugnant person and develop a strong dislike of him. As a result of your antipathy, you focus on the most incriminating evidence against him, resulting in a belief that the defendant is guilty. Rationally speaking, should you revise your belief? It seems to me that you should, even if, holding fixed your antipathy for the defendant, you cannot revise your belief.

From the preceding cases, I conclude that you *can* be rationally required to revise a certain belief even if, holding fixed your other mental states, you are not able to revise that belief. Thus, while rational obligation is constrained by one's current psychological abilities and mechanisms, rational obligation is not constrained by one's current mental states.

Once it is clear that the proposed version of 'ought' implies 'can' does not extend to the view that rational obligations are constrained by one's current *mental states*, it is possible to defend the proposed version against a natural objection, one Sharon Ryan articulates as follows: Suppose you have a much-loved friend who is accused of stealing from you. You believe your friend to be innocent, despite the overwhelming evidence against her. In fact, your affection for your friend is so strong that you are psychologically incapable of revising your belief in your friend's

innocence. According to Ryan, you are rationally obligated to revise your belief in light of the evidence, even if you are psychologically incapable of doing so. Ryan takes this case to refute the claim that epistemic ‘ought’ implies ‘can’ (Ryan 2003, 59).

While Ryan’s case threatens traditional versions of ‘ought’ implies ‘can,’ it leaves unscathed the variant I have proposed. For presumably, what explains your inability to revise your belief in your friend’s innocence is your affection for your friend. But affection is a mental state, not a psychological mechanism or skill. On the variant of ‘ought’ implies ‘can’ proposed, rational obligation is constrained by one’s current psychological mechanisms and skills, but not by one’s current mental states. So, it is not a counter-example to the proposed version of the principle that you might have a rational obligation which your affection for your friend renders you unable to fulfill.

I have been supposing a certain understanding of Ryan’s case, one on which what explains your inability to revise your belief is your affection for your friend. I am supposing, in other words, that were you to lose that affection, you would revise your belief. If this is not the right interpretation of the case—if the persistence of your belief is not explained by your affection for your friend but is rather explained by your lacking certain psychological mechanisms—then the proposed form of ‘ought’ implies ‘can’ does entail that in this case, you are not rationally obligated to revise your belief. But in *this* version of the case, this strikes me as the right result. For if your current repertoire of psychological mechanisms and skills precludes you from revising your belief, it is implausible that you might be rationally obligated to do so. Likewise, it is implausible that the person who cannot read Mandarin might be rationally obligated to revise one of her beliefs in response to a Mandarin text.²⁰

2.4 *The Argument Completed*

The conclusion of the argument is that subjects necessarily have a certain robust psychological ability to revise their beliefs in accordance with relevant counter-evidence. Specifically, subjects are necessarily able to revise their beliefs, given those psychological mechanisms and skills which they already possess. Notably, the claim that subjects are necessarily able to revise their beliefs does not entail that subjects can voluntarily revise their beliefs.

3 **An Objection to the Revisability View: *Idée Fixe***

One natural objection to the revisability view is that the view is too strong. For the revisability view entails that some mental states which we might have thought were beliefs are not beliefs. Specifically, if a subject cannot revise one of her views in the relevant way, then that view is not a belief, even if that subject sincerely endorses that view, relies on it in a wide range of actions, and exploits it as a premise in a wide range of inferences.

For instance, consider the following case: Sheryl forms the view that her neighborhood’s farmer’s market takes place on Fridays. She forms this view in response to excellent evidence. But some point after forming this view, Sheryl suffers a

minor brain lesion which leaves her cognitive faculties entirely intact except for the curious result that she cannot revise her view about the farmer's market. She sees flyers advertising that the market has been rescheduled for Sundays, her friends repeatedly tell her the market is now on Sundays, she even visits the farmer's market on Sundays (quite by accident, since she doesn't anticipate its being held then), but she simply cannot relinquish her view that the market is held on Fridays. She tells anyone who asks that the market is on Fridays. She shows up at the usual spot every Friday, bags and shopping list in tow. She relies on this claim as a premise in a wide range of inferences, such as: *The market is on Fridays. Today is Friday. So, the market is today.* The lesion has transformed Sheryl's prosaic belief into an *idée fixe*. According to the revisability view, Sheryl's *idée fixe* is not a belief, even though it was formed in response to good evidence. It must instead be some other cognitive attitude, such as an entertained thought, an assumption, or a non-doxastic pretense.

The fact that the revisability view excludes *idées fixes* from the class of belief might suggest that the view is too strong. After all, Sheryl sincerely endorses her view about the farmer's market, is guided by it, and relies on it in a wide range of inferences. Since the revisability view excludes mental states like Sheryl's from the class of belief, perhaps we should reject the revisability view in favor of the following view:

THE ANTI-REVISABILITY VIEW OF BELIEF

There exists at least some subject such that: that subject has some belief *p* and has sufficiently strong, undefeated evidence that *not-p*, and that subject is not able, given the psychological mechanisms and skills she already possesses, to revise that belief in accordance with the evidence.

In this section, I defend the result that *idées fixes* are not beliefs. To do this, I first cast doubt on the general theory which I take to support the intuition that *idées fixes* are beliefs. This theory says that if a subject sincerely endorses *p*, is guided in her actions by *p*, and relies on *p* as a premise in a wide range of inferences, then that subject necessarily believes *p*. Put otherwise, sincere assertion, action guidance, and inferential promiscuity are jointly sufficient for belief. I will suggest reasons to doubt that these traits are either individually or jointly sufficient for belief. To be clear, I do not take these considerations alone to refute this theory. Rather, my aim is the more modest one of undermining the theory. Once the theory is in doubt, I then appeal to the motivations for the revisability view to argue that, on balance, we should reject the view that *idées fixes* are beliefs.

First, consider sincere assertion. Perhaps as a general rule, if some subject sincerely asserts *p*, then that subject believes *p*. But, as multiple theorists have argued, this rule seems to have exceptions, most famously in cases of self-deception. Consider the subject who believes that his husband is cheating on him but who also vociferously denies—both to himself and others—that this is the case. Such a subject might sincerely assert that his husband is faithful, while nevertheless exhibiting behavior that is consistent with his belief, such as feeling sad when his husband calls yet again to say that he will be working late or asking his husband more questions than usual about his whereabouts. This sort of case should at least raise doubts

about whether the fact that someone sincerely asserts p guarantees that that person believes p (Cohen 1992, 68–73; Mandelbaum 2014, 79–81).

Second, consider action guidance. Perhaps as a general rule, if some subject is guided in her actions by p , then that subject believes p . But this rule too seems to have exceptions. For mental states other than beliefs, such as pretenses, can also guide action (Gendler 2007, 2008; Velleman 2000). For instance, suppose you are pretending to be an elephant. You might wave your arm in a trunk-like fashion and walk clumsily and slowly. You needn't believe you are an elephant in order for your pretense to govern your actions in this way (Velleman 2000).

In response to the suggestion that pretenses guide action in the same way that beliefs do, it might be replied that pretenses only guide action in a narrow range of contexts, whereas beliefs guide action in a wide range of contexts. While this is an apt reply to cases like the 'elephant' case, the point does not succeed against chronic, far-reaching fantasies, of the sort which plausibly do guide actions across a wide range of circumstances (I will describe such a case shortly and will argue that there is at least some reason to think that the fantasy in question is not a belief.)

Third, consider inferential promiscuity, which is availability as a premise in a wide range of inferences. Perhaps as a general rule, if some subject relies on p in a wide range of inferences, then that subject believes p . However, this rule seems to have exceptions. Specifically, suppositions can also be used as a premise in a wide range of inferences, even when those suppositions are not believed.

Consider the geometry student tasked with proving that no triangle has four sides. To do this, the student might suppose, for the sake of proving otherwise, that there is a triangle which has four sides and then attempt to generate a contradiction from this supposition. We might say that this student 'hypothetically adds to her stock of beliefs' that there is a triangle which has four sides. But whatever it is to hypothetically add to one's stock of beliefs that there is a triangle which has four sides, it is not to believe that there is a triangle that has four sides. It is to suppose it or merely entertain it.

One objection to the view that suppositions are inferentially promiscuous comes from the claim that suppositions cannot contribute to a *wide* range of psychological inferences, whereas beliefs can.²¹ For, the objection goes, while suppositions might figure in inferences involving *reductio*, suppositions do not figure in other sorts of inferences.

Though initially attractive, we should reject the claim that suppositions figure only in a restricted range of inferences. The grain of truth in this claim is that we ordinarily do not make suppositions just for the sake of drawing out their entailments. However, we *can* do this, as we often do in logic and philosophy (and, one might argue, mathematics). For, suppositions have at least as much inferential power as beliefs. Whatever can be believed can be supposed. And whatever inference can be drawn from a set of beliefs can be drawn from a relevantly similar set of suppositions. For instance, I can believe p and believe *if p , then q* and thereby come to believe q . Likewise, I can suppose p and suppose *if p , then q* and thereby come to suppose q .

While it is false that suppositions are not inferentially promiscuous, there is a different trait, one in the vicinity of inferential promiscuity, which likely holds of beliefs but not of suppositions. This is the trait of being usable as a premise in a wide range of psychological inferences, whose conclusions are *believed* (and not merely supposed). While beliefs likely have this trait, suppositions probably do not, since the conclusions of inferences involving suppositions are usually supposed, not believed. For instance, if you suppose, for the sake of argument *if cheese is blue, then Argentina is in Europe*, and you also suppose *cheese is blue*, then you might come to suppose *Argentina is in Europe*, but you probably will not believe that Argentina is in Europe.

Thus, one might suggest the following sufficient condition on belief: If some mental state can serve in a wide range of inferences, whose conclusions are believed, then that state is a belief. This criterion is quite attractive. But notice that it does not suggest a reason to think that *idées fixes* are beliefs. For while it is clear that subjects can rely on *idées fixes* in inferences, it is not clear whether these subjects believe the conclusions of such inferences. It seems at least as likely that such subjects merely suppose the conclusions of such inferences or represent them in some other non-doxastic way.

Suppose, for instance, that Sheryl puts her *idées fixe* to use in the following inference: *The farmer's market is on Fridays. Friday is the day after Thursday. The farmer's market is the day after Thursday.* Does Sheryl believe that the farmer's market takes place the day after Thursday? Or does she merely hold this view in a non-doxastic way, connected as it is to her *idées fixe*? I think the answer is at least unclear. Further, it seems plausible that in order to settle this question, we must settle the question of whether Sheryl's *idées fixe* is a belief. But of course, what we were looking for was some independent way to settle whether Sheryl's *idées fixe* is a belief. Thus, the new proposed condition on belief does not seem particularly useful as a means of determining whether *idées fixes* are beliefs, even if this condition is sufficient for belief.

So far, I've attempted to cast doubt on the view that sincere assertion, action guidance, and inferential promiscuity are individually sufficient for belief. However, it might be claimed that even if these traits do not individually suffice for belief, they jointly suffice for belief. On this view, any mental state which has all three traits is necessarily a belief, whether or not that state is revisable.²²

Perhaps as a general rule, if some subject sincerely endorses *p*, is guided in her actions by *p*, and relies on *p* in a wide range of inferences, then that subject believes *p*. But there are reasons to think this rule has exceptions.

Consider Carolyn, the teenaged protagonist of the novel *Uncle Vampire*. Carolyn develops a detailed and chronic fantasy that her uncle is a vampire. Carolyn sincerely endorses this fantasy, acts on it in a wide variety of circumstances, and relies on it in a wide range of inferences. The fantasy plays a pervasive and important role in her life. This fantasy becomes more intelligible when, in a moment of extreme distress, Carolyn reveals to a therapist that her uncle regularly molests her. It becomes clear that Carolyn's fantasy is some part of an attempt to deal with a traumatic reality (Grant 2015).

The question is, does Carolyn believe that her uncle is a vampire, or does her view take another form? Is it, for instance, a non-doxastic pretense? There are at least two reasons to think that Carolyn does not believe that her uncle is a vampire. First, apart from her ‘vampire’ fantasy, Carolyn is a bright and sociable American teenager. While it is certainly possible that Carolyn believes that vampires exist, it seems at least unlikely that a person of Carolyn’s developmental age and cultural context should harbor such a belief.

Second, on at least one way we might understand Carolyn’s psychology, her ‘vampire’ fantasy is both in conflict with one of her other beliefs and sustained by the very belief with which it is in conflict. For, Carolyn’s fantasy represents that her uncle is a vampire (and not a molester), and Carolyn believes that her uncle is a molester. At the same time, Carolyn’s belief that her uncle is a molester is presumably the cause of her fantasy. For her fantasy is most naturally understood as an attempt to cope with what she believes to be the case.²³

While there is reason to think that beliefs can and do conflict, it is hard to think of another case in which some belief is sustained by the very belief with which it is in conflict. Consider other plausible instances of conflicting belief.²⁴ Perhaps you believe that if you don’t say ‘cancer’ out loud, your partner’s biopsy will come back cancer-free, even whilst believing that there is no connection between your saying ‘cancer’ and your partner’s having cancer (Huddleston 2012). Or, perhaps you believe that there is no external world, having convinced yourself by philosophical argument of this result, whilst also believing—because you cannot help it—that an external world exists.²⁵ In neither of these cases does one of the conflicting beliefs cause or sustain the very belief with which it is in conflict. Moreover, it would be surprising if this were the case. For instance, it would be surprising if your belief that there is no external world should somehow cause or sustain your belief that an external world exists.

It would be unprecedented, then—not to mention odd—for a belief to be sustained by the very belief with which it is in conflict. Since Carolyn’s fantasy would have to be a belief of just this kind, we should be at least doubtful of the suggestion that Carolyn’s fantasy is a belief. It is more natural to think that Carolyn’s fantasy is a non-doxastic pretense, one which is far-reaching and chronic. If this is right, then Carolyn’s fantasy is a counter-example to the view that sincere assertion, action guidance, and inferential promiscuity are jointly sufficient for belief.

I conclude that there are reasons to doubt that sincere assertion, action guidance, and inferential promiscuity are sufficient for belief. So, the fact that *idées fixes* have these traits is not a probative reason to think that such states are beliefs. Moreover, the considerations which motivate the revisability view constitute a reason to think *idées fixes* are not beliefs.

Consider again Sheryl’s *idée fixe* that the farmer’s market is on Fridays. The following two claims—which are instances of the first two premises of the argument from the norm of revision—cannot simultaneously be true of Sheryl:

- (7) Necessarily, if Sheryl believes the farmer’s market is on Fridays and has sufficiently strong, undefeated evidence that the farmer’s market is not on Fridays,

then Sheryl has a *prima facie* obligation of rationality to revise that belief in accordance with that evidence.

- (8) Necessarily, if Sheryl has a *prima facie* obligation of rationality to revise some mental state, then Sheryl can, given those psychological mechanisms and skills which she already possesses, revise that mental state in the required way.

Sheryl has good evidence that the farmer's market is not on Fridays. So, if her view is a belief, then, given (7), she is at least *prima facie* rationally obligated to revise her view. But, if Sheryl has an obligation to revise her view, then, given (8), she must be psychologically capable of revising that view. But, by stipulation, Sheryl lacks this capacity. So, (7) and (8) cannot both be true of Sheryl. However, as I argued in §2, we should accept both of the general principles of which (7) and (8) are instances. So, we have some reason to think that Sheryl's *idée fixe* is not a belief.

Here, then, are two major results about *idées fixes*: First, while *idées fixes* have traits which are also common to beliefs, there are reasons to doubt that such traits are sufficient for belief. Second, the considerations which ground the revisability view provide a reason to think that *idées fixes* are not beliefs. I conclude that, on balance, we should reject the view that *idées fixes* are beliefs. We should posit instead that these states are suppositions, merely entertained thoughts, or non-doxastic pretenses.

4 Summary

I have defended the view that subjects necessarily have a certain robust psychological capacity to revise their beliefs in response to counter-evidence. The argument for this view draws both on a certain norm of rationality and on a variant of 'ought' implies 'can.' The resulting view permits a wide range of mental states into the class of belief, including at least some irrational states and at least some states which are never contravened by evidence. At the same time, the view has the surprising consequence that *idées fixes* are not beliefs. While this result is initially counter-intuitive, I have argued that, on balance, we should welcome it.²⁶

Notes

¹ Those who think subjects are strongly inclined to revise their beliefs in accordance with the evidence include: Adler (2002), Currie and Ravenscroft (2002), Gendler (2008), Shah & Velleman (2005), and Velleman (2000). Those who reject this view include: Bayne & Pacherie (2005, 183), Bortolotti (2012, 39), Gertler (2011), Huddleston (2012), Mandelbaum (2014), and Viedge (forthcoming). In addition, the interpretative view of the mind of the kind associated with Davidson (1984) and Dennett (1989) entails that subjects are necessarily disposed to revise their beliefs in some circumstances, though it is unclear how strong this disposition must be.

² For the view that clinical delusions are not beliefs, see Currie and Jureidini (2001), Currie and Ravenscroft (2002), Frankish (2012), Mullen and Gillett (2014), and Stephens and Graham (2004). For the view that clinical delusions are at least partly beliefs, see Bayne and Pacherie (2005), Bortolotti (2010, 2012), and Radden (2010, 2014). For the view that disavowed prejudices, what are sometimes called *implicit biases*, are not beliefs, see Brownstein and Madva (2016), Levy (2015), Madva (2015), and Schwitzgebel (2010). For the view that disavowed prejudices are beliefs, see Mandelbaum (2016).

For the view that religious convictions are not first-order, factual beliefs, see Rey (2001, 2007) and Van Leeuwen (2014, 2016, 2017). For the view that religious convictions are first-order, factual beliefs, see Boudry and Coyne (2016) and Levy (2017).

³The separate mention of belief suspension is in deference to Friedman's (2013) argument that belief suspension is distinct from both belief elimination and shifts in belief strength. I set aside another way subjects might revise their beliefs, namely by increasing their confidence in those beliefs. I focus on revisions which would be appropriate in response to counter-evidence.

⁴These examples are from or inspired by Chuard & Southwood (2009) and McHugh (2012). See footnote 16.

⁵Likewise, the revisability view can count an omniscient being's views as beliefs, even if this being never possesses counter-evidence for her views (cf. Alston 1986).

⁶If this is right, then norms of rationality are constrained by empirical facts, including—in at least some cases—empirical facts which are discoverable only via experimental psychology. Thus, settling whether some proposed norm of rationality is genuine will sometimes require that we engage with experimental psychology. I welcome this result.

⁷For further discussion, see Helton (2018, 259–260) and Helton (2016, 856–857).

⁸One question about the norm of revision is whether it is *wide-scope* or *narrow-scope*, i.e., whether the obligation to revise one's beliefs is in some sense in effect for subjects who lack relevant counter-evidence, or whether the obligation is not in effect for subjects until they obtain relevant counter-evidence. It is controversial whether norms of rationality in general should be construed as wide-scope or narrow-scope and relatedly, what the implications of these different construals are. See, e.g., Kolodny (2005), cf. Broome (2007). I prefer to let the more general question settle the matter. If it should turn out that norms of rationality are wide-scope, then the norm of revision will also be wide-scope. If it should turn out that norms of rationality are narrow-scope, then the norm of revision will also be narrow-scope. Thanks to Harvey Lederman for discussion on this point.

⁹For a defense of the view that beliefs are not voluntary, see Ryan (2003). For a defense of the view that beliefs are voluntary, see Steup (2008).

¹⁰What I am calling the character-based view of blameworthiness is sometimes termed *attributionism*. The desire-based, value-based, and care-based approaches—which I treat as interchangeable in the text—are in fact distinct and competing approaches. Sources of the character-based view include: Arpaly (2004), Brownstein (2016), Frankfurt (1971), Jaworska (1999, 2007), Sher (2009), Shoemaker (2003, 2011), Smith (2005, 2008, 2012), Sripada (2016), and Watson (1975, 1996).

¹¹Similarly, King (2014) argues that there are obligations to have certain motives and that motives are not under voluntary control. It is at least feasible that those who reject the claim that emotions are governed by obligations, e.g., Pickard (manuscript), might accept my main point via a structurally similar argument from motives.

¹²This now-common strategy of testing an activity's voluntariness is due to Alston (1988, 263).

¹³Here my argument departs from both Chuard and Southwood's and McHugh's, who argue that there are obligations of emotion despite the fact that emotions are *in fact* not under voluntary control (Chuard & Southwood: 2009, 620; McHugh 2012, 87–88). My argument relies instead on the strictly counterfactual claim that if emotions were not under voluntary control, there would still be obligations of emotion.

¹⁴The character-based approach is not the only view which can account for the blameworthiness of involuntary activities. On the *reasons-responsiveness approach*, activities governed by obligations are responsive to reasons in a certain way, whether or not those activities are under voluntary control. See, e.g., Hieronymi (2008, 2014) and McHugh (2012, 89–93; 2013; 2017). (McHugh thinks reasons-responsiveness is not a kind of voluntary control, whereas Hieronymi thinks it is a kind of voluntary control). I have said nothing to exclude the reasons-responsiveness approach, and for present purposes, I do not need to. The character-based and reasons-responsiveness approaches both permit that activities not under voluntary control can be governed by obligations. So, those who prefer the reasons-responsiveness approach can permit me my present point that beliefs can be governed by obligations, whether or not beliefs are under voluntary control.

¹⁵ I am not supposing that someone's being blameworthy entails that it is appropriate for others to blame or otherwise censure that person. See Adams (1985, 21–24), Pickard (2011, 2013, 2014, 2017), and Rosen (2004).

¹⁶ The examples of sneezing and falling asleep are from McHugh (2012, 85, 87). The examples of object discrimination and perception are from Chuard & Southwood (2009, 618).

¹⁷ Glüer and Wikforss (2013a) do not themselves claim that norms must be action-guiding. Their target is action-guiding norms, not norms *per se*.

¹⁸ I thank an anonymous referee for pressing me on this point.

¹⁹ I am indebted to Conor McHugh for discussion on this point and for suggesting a case along these lines.

²⁰ See also McHugh (2012, footnote 62) and Vranas (2007, 183–84). There are, of course, other objections to 'ought' implies 'can', which I cannot do justice to here. See Graham (2011), Mizrahi (2012) and Ryan (2003) for other criticisms of 'ought' implies 'can.' For defenses, see Littlejohn (2012) and Vranas (2007).

²¹ I thank an anonymous referee for raising this concern.

²² I thank both Reina Hayaki and Jake Quilty-Dunn for pressing me on this point.

²³ It might be claimed that Carolyn does not come to believe that her uncle is a molester until the disclosure in the therapist's office and thus, that there is never any conflict between her fantasy and what she believes. This can't be right, for it is presumably Carolyn's awareness that she is being abused which *explains* why she develops the 'vampire' fantasy.

²⁴ Note that the revisability view is consistent with the possibility of conflicting beliefs. The revisability view requires only that subjects can revise their beliefs, not that subjects invariably do revise their beliefs. When this ability is masked, subjects can have conflicting beliefs.

²⁵ This style of example is due to David Chalmers (personal communication). For examples of conflicting mathematical beliefs, see Helton (2015, 90–91) and Mylopoulos (2015, 768).

²⁶ For discussions on this paper, I am indebted to: Martín Abreu Zavaleta, Tom Avery, Michael Bishop, Ned Block, Rosa Cao, David Chalmers, Jeremy Dolan, Daniela Dover, Gabriele Ferretti, Johann Frick, Jane Friedman, Karen Frost-Arnold, Kris Goffin, Laura Gow, Rob Hopkins, Zoe Jenkin, Kyle Landrum, Harvey Lederman, Andrew Lee, Philippe Lusson, Eric Mandelbaum, Conor McHugh, Lisa Miracchi, Bence Nanay, Sam Preston, Jake Quilty-Dunn, Pietro Snider, Maarten Steenhagen, Michael Strevens, Margot Strohminger, Neil Van Leeuwen, Nikolai Viedge, and Nick Young, as well as audiences at: the 2015 Minds Online Conference, New York University, and the University of Nebraska—Lincoln. Special thanks to two anonymous referees.

References

- Adams, R. M. (1985). Involuntary sins. *The Philosophical Review*, 94(1), 3–31.
- Adler, J. (2002). *Belief's own ethics*. Cambridge, MA: MIT Press.
- Alston, W. P. (1986). Does God have beliefs? *Religious Studies*, 22(3–4), 287–306.
- Alston, W. P. (1988). The deontological conception of epistemic justification. *Philosophical Perspectives*, 2, 257–299.
- Arpaly, N. (2004). *Unprincipled virtue: An inquiry into moral agency*. Oxford: Oxford University Press.
- Bayne, T. & Pacherie, E. (2005). In defence of the doxastic conception of delusions. *Mind & Language*, 20(2), 163–188.
- Bortolotti, L. (2010). *Delusions and other irrational beliefs*. Oxford: Oxford University Press.
- Bortolotti, L. (2012). In defence of modest doxasticism about delusions. *Neuroethics*, 5(1), 39–53.
- Boudry, M. & Coyne, J. (2016). Disbelief in belief: On the cognitive status of supernatural beliefs. *Philosophical Psychology*, 29(4), 601–615.
- Broome, J. (2007). Wide or narrow scope? *Mind*, 116(462), 359–370.
- Brownstein, M. (2016). Attributionism and moral responsibility for implicit bias. *Review of Philosophy and Psychology*, 7(4), 765–786.
- Brownstein, M. and Madva, A. (2012). The Normativity of automaticity. *Mind & Language*, 27(4), 410–434.
- Card, C. (1988). Gratitude and obligation. *American Philosophical Quarterly*, 25(2), 115–127.

- Chuard, P. & Southwood, N. (2009). Epistemic norms without voluntary control. *Noûs*, 43(4), 599–632.
- Cohen, L. J. (1992). *An essay on belief and acceptance*. Oxford: Clarendon Press.
- Currie, G. & Jureidini, J. (2001). Delusion, rationality, empathy: commentary on Martin Davies et al. *Philosophy, Psychiatry, & Psychology*, 8(2), 159–162.
- Currie, G. & Ravenscroft, I. (2002). *Recreative minds*. Oxford: Oxford University Press.
- Davidson, D. (1984). *Inquiries into truth and interpretation*. Clarendon: Oxford University.
- Dennett, D. C. (1989). *The intentional stance*. Cambridge, MA: MIT Press.
- Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *The Journal of Philosophy*, 68(1), 5–20.
- Frankish, K. (2012). Delusions, levels of belief, and non-doxastic acceptances. *Neuroethics*, 5(1), 23–27.
- Friedman, J. (2013). Suspended judgment. *Philosophical Studies*, 162(2), 165–181.
- Gendler, T. S. (2007). Self-deception as pretense. *Philosophical Perspectives*, 21(1), 231–258.
- Gendler, T. S. (2008). Alief and belief. *The Journal of Philosophy*, 105(10), 634–663.
- Gertler, B. (2011). Self-knowledge and the transparency of belief. *Self-Knowledge*, Ed. A. Hatzimoysis, 125–145. Oxford: Oxford University Press.
- Glüer, K. & Wikforss, Å. (2013a). Against belief normativity. *The Aim of Belief*, ed. T. Chan, 80–99. Oxford: Oxford University Press.
- Glüer, K. & Wikforss, Å. (2013b). Aiming at truth: on the role of belief. *Teorema*, 32(3), 137–162.
- Graham, P. A. (2011). ‘Ought’ and ability. *The Philosophical Review*, 120(3), 337–382.
- Grant, C. D. (2015). *Uncle vampire*. Open Road Media.
- Harmon-Jones, E. (2000). A cognitive dissonance theory perspective on the role of emotion in the maintenance and change of beliefs and attitudes. *Emotions and Beliefs*, Eds. N.H. Frijada, A.R.S. Manstead, and S. Bem, 185–211. Cambridge, England: Cambridge University Press.
- Harmon-Jones, E. & Harmon-Jones, C. (2007). Cognitive dissonance theory after 50 years of development. *Zeitschrift für Sozialpsychologie*, 38(1), 7–16.
- Helton, G. (2015). *Two worlds, one mind: The divide between perception and belief*. (unpublished doctoral dissertation). New York: New York University.
- Helton, G. (2016). Recent issues in high-level perception. *Philosophy Compass*, 11(12), 851–862.
- Helton, G. (2018). Visually perceiving the intentions of others. *The Philosophical Quarterly*, 68(271), 243–264.
- Hieronymi, P. (2008). Responsibility for believing. *Synthese*, 161(3), 357–373.
- Hieronymi, P. (2014). Reflection and responsibility. *Philosophy & Public Affairs*, 42(1), 3–41.
- Huddleston, A. (2012). Naughty beliefs. *Philosophical Studies*, 160(2), 209–222.
- Jaworska, A. (1999). Respecting the margins of agency: Alzheimer’s patients and the capacity to value. *Philosophy & Public Affairs*, 28(2), 105–138.
- Jaworska, A. (2007). Caring and internality. *Philosophy and Phenomenological Research*, 74(3), 529–568.
- King, A. (2014). Actions that we ought, but can’t. *Ratio*, 27(3), 316–327.
- Kolodny, N. (2005). Why be rational? *Mind*, 114(455), 509–563.
- Levy, N. (2015). Neither fish nor fowl: Implicit attitudes as patchy endorsements. *Noûs*, 49(4), 800–823.
- Levy, N. (2017). Religious beliefs are factual beliefs: Content does not correlate with context sensitivity. *Cognition*, 161, 109–116.
- Littlejohn, C. (2012). Does ‘ought’ still imply ‘can’? *Philosophia*, 40(4), 821–828.
- Madva, A. (2016). Why implicit attitudes are (probably) not beliefs. *Synthese*, 193(8), 2659–2684.
- Mandelbaum, E. (2014). Thinking is believing. *Inquiry*, 57(1), 55–96.
- Mandelbaum, E. (2016). Attitude, inference, association: On the propositional structure of implicit bias. *Noûs*, 50(3), 629–658.
- McHugh, C. (2012). Epistemic deontology and voluntariness. *Erkenntnis*, 77(1), 65–94.
- McHugh, C. (2013). Epistemic responsibility and doxastic agency. *Philosophical Issues*, 23(1), 132–157.
- McHugh, C. (2017). Attitudinal control. *Synthese*, 194(8), 2745–2762.
- Mizrahi, M. (2012). Does ‘ought’ imply ‘can’ from an epistemic point of view? *Philosophia*, 40(4), 829–840.
- Mullen, R. & Gillett, G. (2014). Delusions: A different kind of belief? *Philosophy, Psychiatry, & Psychology*, 21(1), 27–37.

- Mylopoulos, M. I. (2015). Agentive awareness is not sensory awareness. *Philosophical Studies*, 172(3), 761–780.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2): 175–220.
- Pickard, H. (2011). Responsibility without blame: Empathy and the effective treatment of personality disorder. *Philosophy, Psychiatry, & Psychology*, 18(3), 209–223.
- Pickard, H. (2013). Responsibility without blame: Philosophical reflections on clinical practice. *Oxford Handbook of Philosophy of Psychiatry*, 1134–1154.
- Pickard, H. (2014). Responsibility without blame: Therapy, philosophy, law. *Prison service journal*, 213, 10–16.
- Pickard, H. (2017). Responsibility without blame for addiction. *Neuroethics*, 10(1), 169–180.
- Pickard, H. (manuscript). Stop telling me how to feel! A clinical theory of emotions and what's wrong with the moralization of feelings.
- Radden, J. (2010). *On Delusion*. London: Routledge.
- Radden, J. (2014). Belief as delusional and delusion as belief. *Philosophy, Psychiatry, & Psychology*, 21(1), 43–46.
- Rey, G. (2001). Meta-atheism: Reasons for thinking few people actually believe in god. *The Experience of Philosophy*, Eds. R. Martin and D. Kolak. Oxford University Press.
- Rey, G. (2007). Meta-atheism: religious avowal as self-deception, *Philosophers without Gods: Meditations on Atheism and the Secular Life*, 243–265, Ed. L.M. Antony, New York: Oxford University Press.
- Rosen, G. (2004). Skepticism about moral responsibility. *Philosophical Perspectives*, 18(1), 295–313.
- Rusin, J. (2010). Reconciling epistemic duty with non-voluntary belief. *International Journal of the Humanities*, 8(8), 95–101.
- Ryan, S. (2003). Doxastic compatibilism and the ethics of belief. *Philosophical Studies*, 114(1), 47–79.
- Schwind, C., Buder, J., Cress, U. & Hesse, F. W. (2012). Preference-inconsistent recommendations: An effective approach for reducing confirmation bias and stimulating divergent thinking? *Computers & Education* 58(2), 787–796.
- Schwitzgebel, E. (2001). In-between believing. *The Philosophical Quarterly*, 51(202), 76–82.
- Schwitzgebel, E. (2002). A dispositional, phenomenal account of belief. *Noûs*, 36(2), 249–275.
- Schwitzgebel, E. (2010). Acting contrary to our professed beliefs or the gulf between occurrent judgment and dispositional belief. *Pacific Philosophical Quarterly*, 91(4), 531–553.
- Shah, N. & Velleman, J. D. (2005). Doxastic deliberation. *The Philosophical Review*, 114(4), 497–534.
- Sher, G. (2009). *Who knew? Responsibility without awareness*. Oxford: Oxford University Press.
- Shoemaker, D. (2003). Caring, identification, and agency. *Ethics*, 114(1), 88–118.
- Shoemaker, D. (2011). Attributability, answerability, and accountability: Toward a wider theory of moral responsibility. *Ethics*, 121(3), 602–632.
- Smith, A. (2005). Responsibility for attitudes: Activity and passivity in mental life. *Ethics*, 115(2), 236–271.
- Smith, A. (2008). Control, responsibility, and moral assessment. *Philosophical Studies*, 138, 367–392.
- Smith, A. (2012). Attributability, answerability, and accountability: In defense of a unified account. *Ethics*, 122(3), 575–589.
- Sripada, C. (2016). Self-expression: A deep self theory of moral responsibility. *Philosophical Studies*, 173(5), 1203–1232.
- Stephens, G. L. & Graham, G. (2004). Reconceiving delusion. *International Review of Psychiatry*, 16(3), 236–241.
- Steup, M. (2008). Doxastic freedom. *Synthese*, 161(3), 375–392.
- Stich, S. P. (1978). Beliefs and subdoxastic states. *Philosophy of Science*, 499–518.
- Svedholm-Häkkinen, A. M. (2015). Highly reflective reasoners show no signs of belief inhibition. *Acta psychologica*, 154, 69–76.
- Thompson, V. A. & Johnson, S. C. (2014). Conflict, metacognition, and analytic thinking. *Thinking & Reasoning*, 20(2), 215–244.
- Viedge, N. (forthcoming). Defending evidence-resistant beliefs. *Pacific Philosophical Quarterly*.
- Van Leeuwen, N. (2014). Religious credence is not factual belief. *Cognition*, 133(3), 698–715.

- Van Leeuwen, N. (2016). Beyond fakers and fanatics: A reply. *Philosophical Psychology*, 29(4), 616–621.
- Van Leeuwen, N. (2017). Two paradigms for religious representation: The physicist and the playground (a reply to Levy). *Cognition*, 164, 206–211.
- Velleman, D. (2000). *The possibility of practical reason*. Oxford: Oxford University Press.
- Vranas, P. (2007). I ought, therefore I can. *Philosophical Studies* 136(2), 167–216.
- Watson, G. (1975). Free agency. *Journal of Philosophy* 72, 205–220.
- Watson, G. (1996). Two faces of responsibility. *Philosophical Topics* 24(2), 227–248.