

Belief as a Feeling of Conviction

Declan Smithies

Forthcoming in *The Nature of Belief*,
edited by E. Schwitzgebel and J. Jong (Oxford University Press)

What is belief? My view is that belief is a *feeling of conviction*: to believe that p is to feel – or to be disposed to feel – convinced that p . I've argued before that feeling conviction is *necessary* for belief: if you believe that p , then you're disposed to feel convinced that p (Smithies 2012a & 2019: Ch. 4). My goal here is to argue that feeling conviction is *sufficient* for belief: if you're disposed to feel conviction that p , then you believe that p .

In §1, I outline a *normative constraint* on theories of belief. There are doxastic norms of epistemic and practical rationality that apply to all beliefs as such. One task for a theory of belief is to articulate the essential nature of the state that is subject to these norms of rationality.

In §2, I consider the functionalist view that the essence of belief is to play a certain causal role – namely, one that approximately satisfies doxastic norms of rationality. I argue that functionalism confronts a version of the problem of chauvinism that it was originally designed to avoid, since it cannot explain the conceivability of a *madman* whose beliefs play dysfunctional causal roles that systematically violate doxastic norms of rationality.

In §3, I explain how a phenomenal conception of belief avoids the problem of chauvinism. On this view, the essence of belief is to feel convinced that what you believe is true. A madman's beliefs are feelings of conviction that play dysfunctional and irrational causal roles.

In §4, I address the objection that only dispositional states are beliefs. I argue that occurrent feelings of conviction and standing dispositions to feel conviction should be regarded as two distinct ways of believing, since they are subject to the same doxastic norms of rationality that apply to all beliefs as such. On this view, occurrent and dispositional belief are two distinct determinates of a common determinable.

In §5, I address the objection that feelings of conviction are not sufficient for belief unless they play an appropriate functional role. Again, my response appeals to the normative

characterization of belief as a state that is subject to doxastic norms of rationality. All feelings of conviction are subject to doxastic norms of rationality whether or not they conform to them. Hence, all feelings of conviction are beliefs.

In §6, I address the objection that my view generates problematic results in cases of implicit bias (Schwitzgebel 2010, 2021). I respond that we must attribute egalitarian beliefs to subjects who display implicit bias in order to hold them accountable for violating the doxastic norms of rationality that apply to all beliefs as such.

Finally, in §7, I end with some concluding methodological reflections.

1. The Normative Profile of Belief

What are the ground rules for inquiring into the nature of belief? We need some guidance to avoid purely terminological debates in which different theorists use the word ‘belief’ to pick out distinct mental states that are equally real and theoretically important. David Chalmers offers the following sensible advice:

Instead of asking ‘What is X?’ one should focus on the roles one wants X to play and see what can play that role. (2011: 538)

The suggestion is that any inquiry into the nature of belief must be constrained by some antecedent specification of which theoretical roles it is supposed to play. Only then can we ask what belief must be like in order to play these theoretical roles. Otherwise, disagreements about the nature of belief risk degenerating into purely verbal debates.

So let’s begin by specifying which theoretical roles we associate with the concept of belief. Why do we attribute beliefs to one another in the first place? One familiar answer is that we use the concept of belief in *prediction* and *explanation*. This is perhaps the central motivation behind functionalist theories of mind. According to David Lewis, for example, our concepts of belief and desire are implicitly defined by their role in a psychological theory – namely, *folk psychology* – that we use to predict and explain people’s behavior based on their beliefs and desires.

According to Lewis, folk psychology is a causal theory of mind. In this respect, it is no different from other causal theories, such as folk physics, which we use to predict and explain the movements of physical objects. Nevertheless, folk psychology has distinctive features that set it apart from other causal theories: most importantly, it is not just a *causal* theory, but a *normative* one. We use our concepts of belief and desire not only to predict and explain behavior, but also to evaluate whether such behavior complies with norms of rationality. These normative evaluations play an important role in our social lives. We hold each other accountable for meeting normative standards of rationality in our beliefs, desires, and actions, and when we violate these standards, we subject one another to reactive attitudes, including blame and resentment. Just as folk psychology contains causal information that we use in prediction and explanation, so it also contains normative information that we use in evaluation.

Although it is often neglected, Lewis explicitly recognizes this point. For example, he says that normative theories of rationality, such as decision theory and probability theory, are “severely idealized parts of folk psychology” (1999: 321). Moreover, the causal and normative dimensions of folk psychology are not completely independent of each other. According to Lewis, the causal roles that beliefs and desires play in responding to sensory inputs, producing behavioral outputs, and interacting with each other are approximately similar to the roles they rationally *should* play. As Lewis writes: “Folk psychology says that we make sense. It credits us with a modicum of rationality in our acting, believing, and desiring” (1999: 320).

Lewis is not idiosyncratic in this respect. Mainstream functionalist theories of mind typically presuppose some degree of rationality in specifying the causal roles of belief and desire. Consider, for instance, Robert Stalnaker’s *pragmatic picture* of mind:

Belief and desire . . . are correlative dispositional states of a potentially rational agent. To desire that p is to be disposed to act in ways that would tend to bring it about that p in a world in which one’s beliefs, whatever they are, were true. To believe that p is to be disposed to act in ways that would tend to satisfy one’s desires, whatever they are, in a world in which p (together with one’s other beliefs) were true. (1984: 15)

The implicit suggestion here is that the causal role of belief and desire approximates its normative role in making action rational. Moreover, this appeal to rationality seems unavoidable in specifying the causal roles of belief and desire. After all, it is at least conceivable that there can be perfectly rational agents – such as gods or angels – who have beliefs and desires. If so, then no irrational causal role is required for belief or desire. And it is implausible that any non-rational causal role is required for belief or desire, rather than merely contingently associated with them. Hence, any causal role required for belief and desire must be rational (Wedgwood 2007).

The main goal of this chapter is to argue that we should abandon functionalism as a theory of belief and replace it with the phenomenal theory of belief as a feeling of conviction. But my preliminary goal in this section is to find some common ground on which all participants to this debate can agree. I propose that the normative profile of belief can serve as a neutral starting point for debates about the nature of belief. As I've explained, mainstream functionalists recognize that our concept of belief has a normative dimension as well as a causal one. Hence, both sides can agree that belief is subject to distinctive norms of rationality. The disagreement concerns the nature of the state that is subject to these rational norms. Does its essence consist in its causal role or its phenomenal character?

Of course, we can disagree about which norms of rationality govern belief. For instance, epistemologists debate whether rationality requires your beliefs to be logically consistent and closed under logical consequence or whether it imposes only the weaker requirement of probabilistic coherence. Without settling these debates, we can agree that there are distinctive norms of rationality that uniquely apply to beliefs as distinct from other mental states. To a first approximation, all beliefs are subject to rational norms of three different kinds:

- (1) *Input norms*: Your beliefs should cohere with perceptual inputs, e.g. you should believe that things are how they perceptually appear in the absence of defeaters.
- (2) *Output norms*: Your beliefs should cohere with behavioral outputs given your desires, e.g. you should act in ways that would satisfy your desires given what you believe.
- (3) *Interaction norms*: Your beliefs should cohere with other beliefs, e.g. your beliefs should be either logically or probabilistically coherent.

Formal theories of rationality, including probability theory and decision theory, attempt to articulate these rational norms more precisely. For our purposes, however, no such precision is required. It suffices to acknowledge that there are distinctive norms of rationality that apply uniquely to beliefs as such. Other mental states, including desires, intentions, and emotions, may be subject to their own proprietary norms of rationality. But the rational norms that govern belief are distinct from those that govern other mental states. Hence, beliefs can be distinguished from other mental states by the doxastic norms of rationality that apply uniquely to them.

This is not yet to distinguish between *binary* and *graded* concepts of belief. I'll focus on the binary conception of belief as an all-or-nothing state, rather than the graded conception of belief as a state that comes in varying degrees of confidence. Graded and binary belief may be regarded as species of a common genus insofar as their inputs, outputs, and mutual interactions are governed by overlapping norms of rationality. Even so, they are distinct species of the genus because there are distinctive norms governing binary belief that do not apply to graded belief. Arguably, binary belief is subject to a *knowledge norm*, which says that you shouldn't believe what you don't know (Williamson 2000; Smithies 2012b). It is irrational to believe something in the binary sense when you know you don't know it, whereas there is no inherent irrationality in having high confidence about something you know you don't know. Thus, binary belief is distinct from high confidence. In this chapter, I'll follow the usual practice of using the term 'belief' exclusively in its binary sense.

I suggest that we can avoid terminological debates about belief by first giving a neutral characterization in terms of its normative profile. To believe something is to be in a state that is subject to doxastic norms of rationality that apply to all beliefs as such. We can agree on this much without settling on any precise formulation of exactly what those norms are.

These are two points to distinguish here. First, all beliefs are subject to doxastic norms of rationality. Not all belief-like representational states are subject to the same doxastic norms as our beliefs. For instance, our understanding of the syntax of our language may be explained by our implicit representation of syntactic principles, but there is no rational requirement that these implicit representations should cohere with our beliefs. Since they are not subject to the same

rational norms of coherence that govern our beliefs, these representational states are not beliefs at all. Instead, they are what Stephen Stich (1978) calls *subdoxastic* states.

Second, all states that are subject to doxastic norms of rationality are beliefs. Of course, we can divide beliefs into further subcategories depending on their subject matter, physical realization, or causal role. So, for example, we can distinguish scientific beliefs from mathematical beliefs, human beliefs from Martian beliefs, rational beliefs from irrational beliefs, and so on. However, it would be *chauvinistic* to deny that a state is a genuine belief on any such basis when it is subject to the same rational norms as other beliefs. One of the main goals of this chapter is to argue that functionalist theories of belief are guilty of chauvinism.

In summary, belief can be characterized in normative terms as a state that is subject to doxastic norms. However, this normative characterization of belief is not fundamental. We need some non-normative account of the essence of belief that explains why it is subject to doxastic norms. After all, the norms governing belief are distinct from the norms governing other mental states, including desires, intentions, and emotions. Presumably, this is not just a brute fact that resists any further explanation. There must be some non-normative difference between these mental states that explains why they are subject to different norms of rationality.

This much should be common ground between competing theories of the nature of belief. No one should deny that belief is a state that is subject to certain doxastic norms of rationality that apply to all beliefs as such. The real dispute concerns the nature of the state that is subject to these doxastic norms. To make progress in resolving this dispute, we need to ask what beliefs must be like in order to be subject to doxastic norms of rationality. This is a substantive question and not a merely terminological one.

2. Functionalism and the Problem of Chauvinism

Functionalism is the thesis that the essence of belief is to play a certain causal role. There is no conflict here with the claim that belief is also subject to doxastic norms of rationality. Instead, functionalism offers a more fundamental characterization of belief that explains why it is subject to doxastic norms. According to functionalism, beliefs and desires are subject to different norms

of rationality because they play different causal roles. It is in virtue of playing a belief-like causal role, rather than a desire-like causal role, that beliefs are subject to doxastic norms of rationality.

As noted in §1, the causal role of belief is usually specified explicitly or implicitly using the concept of rationality. According to Lewis's version of functionalism, for instance, belief is a state that plays a causal role that is approximately rational in the sense that it comes close enough to satisfying the doxastic norms that apply to beliefs as such. On this view, beliefs are governed by doxastic norms because they have some tendency to conform to them. As a result, Lewis's functionalism imposes *rationality constraints* on belief. These constraints are weak enough to allow for some degree of irrationality, since beliefs need only be rational enough to meet some minimal standard. If a state has no tendency whatsoever to conform to doxastic norms of rationality, however, then it is not a belief at all. Hence, there are limits on how much irrationality is compatible with having beliefs.

As I've argued elsewhere, it is dubious that there are any such rationality constraints on belief (Smithies, Lennon, & Samuels 2022). There is substantial empirical evidence of human irrationality from the psychology of reasoning and the psychopathology of delusion. Moreover, it seems conceivable that there could be Lewisian madmen whose beliefs are even more grossly irrational than we find in human psychology. To exclude this possibility is to confront a version of the *problem of chauvinism* that functionalism was originally designed to avoid. Functionalism seeks to avoid chauvinism by allowing that beliefs can be multiply realized in different kinds of physical systems. For instance, we can share our beliefs with Martians whose alien physiology plays the requisite causal role. And yet the problem of chauvinism remains unless we can equally accommodate madmen whose beliefs play abnormal and irrational causal roles.

Lewis takes this problem very seriously. For instance, he explicitly acknowledges the possibility of mad pain:

There might be a strange man who sometimes feels pain, just as we do, but whose pain differs greatly from ours in its causes and effects. (1983: 122)

Moreover, he recognizes that the same possibility arises for propositional attitudes, such as belief and desire, which can also play abnormal and irrational causal roles:

The same possibility should be recognized for attitudes as well. Karl might believe himself a fool, and might desire fame, even though the best interpretation of Karl considered in isolation might not assign those attitudes to him. (1983: 119)

For current purposes, it doesn't much matter whether these scenarios are genuinely possible, as Lewis maintains, or whether they are merely conceivable. After all, analytic functionalism builds rationality constraints into the analysis of our ordinary concept of belief. How then can Lewis explain the conceivability of mad belief?

His solution is that a madman can have beliefs and desires by having physical states of the same kind that normally play a belief-like causal role in us or in others of his kind. So we can imagine a *madman* whose beliefs are physically similar but functionally different from our own. We can even imagine a *mad Martian* whose beliefs are both physically and functionally dissimilar from ours so long as they are physically similar to Martian beliefs that normally function like ours. What Lewis cannot explain is our ability to imagine an entire *population* of mad Martians whose beliefs are both physically and functionally dissimilar from ours. In the limit case, for instance, we can imagine a *solitary mad Martian* in a population of one. Lewis is forced to deny that a solitary mad Martian can have beliefs, although he allows that its intrinsic duplicates may have beliefs when they are members of an appropriate population. Lewis concedes that this result is highly counterintuitive, but he sees no viable alternative for functionalism:

Any broadly functionalist theory of mind is under intuitive pressure from two directions. On the one hand, it seems wrong to make it invariable or necessary that the mental states occupy their definitive causal roles. Couldn't there be occasional exceptions . . .? On the other hand, the mental states of Karl seem intrinsic to him. Why should whether he now feels pain – or believes himself to be a fool, or desires fame – depend on what causes

what in the case of someone else? I do not see any acceptable way to respect both intuitions in their full strength. (1983: 120)

To my mind, this amounts to a concession that functionalism cannot adequately solve the problem of chauvinism. Moreover, functionalism is not the only viable option. As I'll argue in §3, there is a better alternative.

Can functionalists avoid the problem of chauvinism by abandoning Lewis's commitment to rationality constraints on belief? This is easier said than done, since it's not clear how to specify the functional role of belief without appealing implicitly or explicitly to the concept of rationality. In any case, the problem of chauvinism remains. After all, it is conceivable that a madman can have beliefs that play abnormal functional roles. So the problem of chauvinism cannot be solved merely by specifying the functional role of belief without using the concept of rationality.

As far as I can see, the problem of chauvinism plagues all versions of functionalism. It afflicts not only commonsense functionalism, which specifies the functional role of belief in terms of folk psychology, but also psycho-functionalism, which specifies its functional role in terms of scientific psychology (Block 1978). It arises for functional role theories of propositional attitudes (Fodor 1987) and contents (Harman 1987) as well as theories that encompass both (Lewis 1999). It threatens theories that prioritize specific aspects of the functional role of belief, including its role in inference (Stich 1978), action (Stalnaker 1984), assertion (Kaplan 1996), responsiveness to evidence (Velleman 2000), or rational revisability (Helton 2020). And, most importantly for current purposes, it arises for semi-functional theories on which playing the right causal role is necessary but insufficient for belief (Schwitzgebel 2002).

One goal of this chapter is to explain why functionalism cannot escape the problem of chauvinism simply by biting the bullet – that is, by denying that mad belief is conceivable at all (Schwitzgebel 2012). In earlier work, I've argued against this move by highlighting the continuity between mad belief and delusional belief (Smithies, Lennon, & Samuels 2022). In this chapter, I adopt a different but complementary strategy. I argue that we cannot adequately explain why delusional subjects and madmen are irrational unless we recognize that their beliefs violate doxastic norms of rationality. Ultimately, however, I suspect that the tendency to deny the

conceivability of mad belief derives from the false impression that functionalism is the only viable option for a theory of propositional attitudes. So the next section presents an alternative theory of belief that avoids the problem of chauvinism.

3. The Feeling of Conviction

How can we avoid the problem of chauvinism by explaining the conceivability of mental states in Martians and madmen? My answer is that we have *phenomenal concepts* of mind that cannot be analyzed either in terms of physical realization or functional role.

Let's begin with the concept of *pain*. The feeling of pain has a distinctive phenomenal character: there is something it's like to feel pain, which is different from what it's like to feel hunger, desire, or anything else. Moreover, we know by *introspection* what it's like to feel pain: in other words, we know about our own experience in a distinctively first-personal way, which is different from our third-personal ways of knowing about the experiences of other people. Introspection enables us to know what it's like to feel pain without knowing anything about its physical realization or functional role. Moreover, our introspective knowledge of what it's like to feel pain enables us to understand what it's like for *someone else* to feel pain – that is, to have an experience that feels *just like this*. Thus, introspection gives us first-personal ways of thinking about our own experience – that is, *phenomenal concepts* (Chalmers 1996).

Our possession of phenomenal concepts explains how we can imagine Martians and madmen whose mental states feel just like ours. When we imagine the feeling of pain using a phenomenal concept, we imagine a state that feels just like our own pain. And we can imagine this without making any further assumptions about its physical realization or functional role. This enables us to imagine scenarios in which the feeling of pain is dissociated from its normal physical realization or functional role. That is how we can make sense of the hypothesis that a solitary mad Martian feels pain, just as we do, although its feelings diverge from ours both in physical realization and functional role.

I contend that we can imagine mad belief in much the same way as mad pain. We have a *phenomenal concept* of belief as a feeling of conviction. After all, belief has a phenomenal dimension: there is something it's like to feel convinced that something is true. The feeling of

conviction has its own distinctive phenomenal character, which distinguishes it not only from sensory feelings of pain or hunger, but also from cognitive feelings of uncertainty or doubt. Moreover, we know by introspection what it's like to experience this feeling of conviction. And we can use this knowledge to understand what it's like for someone else to feel the same way. We can imagine someone who feels conviction, just as we do, although their feelings of conviction diverge from ours in both physical realization and functional role. This is how we can make sense of the hypothesis that a solitary mad Martian has beliefs.

Here is an example to illustrate I mean by the feeling of conviction. Before the Presidential Election in November 2020, I recall feeling initially doubtful that Biden would beat Trump. As the votes started to come in, I began to feel increasingly confident that Biden would be elected, although the race was still too close to make any definitive judgment call. When the decisive results came in a few days later, however, I felt convinced that Biden was going to win after all. Although the outcome would not be officially confirmed until several weeks later, it now seemed to me that the result was settled beyond any reasonable doubt. I felt as if I knew the outcome.

In previous work, I used the term 'judgment' to refer to these feelings of conviction (Smithies 2012a & 2019: Ch. 4). The same term is sometimes used in a more restricted sense to refer to the mental act of deliberately making up your mind based on conscious reflection on the evidence (Peacocke 2007). This is one way of feeling conviction, but it's not the only way. Feelings of conviction may occur unbidden without any mental activity of conscious reflection, deliberation, or making up your mind. For instance, it might just suddenly occur to you that you left your wallet at home.

My account relies on three background assumptions. First, the feeling of conviction is a *cognitive* feeling, rather than a *sensory* feeling. I see no reason to accept any form of empiricist reductionism, according to which all experience must be reducible to sensory perception construed broadly to include bodily sensations, sensory imagery, and inner speech. In addition to sensory feelings, we also have cognitive feelings of thinking, understanding, and reasoning as well as confidence, doubt, and conviction. There is no pressure to reductively explain these cognitive feelings in terms of bodily sensations, sensory imagery, or inner speech.

Second, feeling conviction has an *attitude-specific* phenomenal character: what it's like to feel convinced that p is different from what it's like to adopt any other attitude towards the same proposition, such as feeling confident or doubtful that p . There is a phenomenal contrast between feeling initially doubtful that p , feeling increasingly confident that p , and eventually feeling convinced that p . I see no way to explain these phenomenal contrasts in more basic terms. The feeling of conviction is just that – a feeling of *conviction*, rather than confidence or doubt.

Third, feeling conviction has a *content-specific* phenomenal character: for instance, what it's like to feel convinced that p is different from what it's like to feel convinced that not- p or to feel convinced that p given the assumption that q . Again, there is a phenomenal contrast between initially feeling convinced that p , then retreating to the feeling of conviction that p given that q , and finally feeling convinced that neither p nor q is true. Feelings of conviction are not *cognitive qualia* that can be experienced without any intentional content. To feel conviction is to feel conviction *that something is so*. Moreover, the phenomenal character of feeling conviction varies depending on what you feel convinced is true. No doubt, phenomenal duplicates can think thoughts with different contents in virtue of differences in their external environment. When external circumstances are held fixed, however, what it's like to feel conviction about one thing is different from what it's like to feel conviction about anything else.

I won't defend these assumptions here, since they are extensively defended elsewhere (see Smithies 2013a & 2013b for an overview). Instead, my goal is to engage with an opponent who is willing to grant these assumptions – at least for the sake of argument – but who denies that feelings of conviction are sufficient for belief. This opponent concedes that we can imagine madmen who experience feelings of conviction as well as feelings of pain. According to my opponent, however, this is not sufficient for the madman to have beliefs. In the next two sections, I'll consider two versions of this objection.

4. Occurrent and Dispositional Belief

The first objection is that belief is a *dispositional state*, whereas judgments or feelings of conviction are *occurrent episodes* (Crane 2013). On this view, it's a category mistake to regard feelings of conviction as beliefs, since they are not dispositional states.

The grain of truth in this objection is that we do have a dispositional concept of belief. Your beliefs about something can persist over time even when you're thinking about something else or nothing at all. These standing beliefs are associated with dispositions that need not be manifested at any given time. For instance, you can believe that p without currently feeling convinced that p . Hence, the occurrent feeling of conviction is not necessary for belief.

Nevertheless, our standing beliefs are normally disposed to manifest themselves in phenomenal consciousness as occurrent feelings of conviction, although these dispositions can be masked by interfering factors, such as repression or self-deception. Arguably, some disposition to feel conviction that p is *necessary* for believing that p in the dispositional sense (Smithies 2012a & 2019: Ch. 4). All I need for current purposes, however, is that the disposition to feel conviction that p is *sufficient* for believing that p in the dispositional sense. So, for example, if the madman is disposed to feel convinced that p , then he believes that p .

Although I've disputed this in the past (Smithies 2019: 50), I'm now persuaded that we have an *occurrent* concept of belief as well as a *dispositional* one. We can recognize the distinction between occurrent and dispositional belief without subscribing to a dubious metaphysical picture on which beliefs move in and out of consciousness. As I've insisted before, dispositional beliefs are distinct from their occurrent manifestations. Even so, we can acknowledge this distinction without restricting the concept of belief to dispositional states as opposed to their occurrent manifestations. Instead, we should regard occurrent and dispositional belief as distinct determinates of a common determinable: these are just two different ways of believing a proposition.

I'll now argue for this claim by appealing to the normative profile of belief outlined in §1. To believe something is to be in a state that is subject to doxastic norms of rationality that apply to all beliefs as such. These doxastic norms apply not only to our standing dispositions, but also to our occurrent feelings of conviction. That is why it's irrational to feel convinced that something is true when it is inconsistent with all your evidence and with everything else you believe. Hence, we should recognize that we have occurrent beliefs as well as dispositional beliefs.

In previous work, I described the occurrent manifestations of belief in the stream of phenomenal consciousness as *judgments*, rather than *beliefs* (Smithies 2019: 131). But it now

seems to me spurious to deny that judgments are beliefs once we grant that they are subject to the same doxastic norms. Instead, we should say that judgment is an occurrent episode of belief, since it's subject to the same doxastic norms that apply to all beliefs as such. I previously dismissed this view on the grounds that slips in judgment need not express what you believe (Smithies 2019: 184). But these cases are better described as ones in which your judgments are fleeting occurrent beliefs that conflict with your more stable dispositional beliefs. This description better captures the irrationality of the situation: you violate doxastic norms of rationality insofar as there is some temporary conflict within your belief system.

Someone might protest that occurrent feelings of conviction are not subject to all the same rational norms as standing dispositions to feel conviction. For instance, an occurrent feeling of conviction makes it rational to believe that you now feel conviction, whereas a mere disposition to feel conviction does no such thing. In reply, however, this only shows that occurrent and dispositional belief are *distinct determinates* of a common determinable. Even so, there is enough overlap in the norms governing occurrent and dispositional belief to regard them as distinct determinates of a *common determinable*. Your beliefs – whether dispositional or occurrent – should cohere in the right ways with each other, with perceptual inputs, and with behavioral outputs. Moreover, your first-order beliefs – whether dispositional or occurrent – should cohere with your higher-order beliefs about what you believe. If we deny that occurrent feelings of conviction are beliefs at all, then we obscure this normative overlap with our dispositional beliefs.

In conclusion, it is no category mistake to regard feelings of conviction as beliefs, since we have an occurrent concept of belief as well as a dispositional one. I contend that feeling conviction is sufficient for belief in this occurrent sense. Feeling convinced that p is a way of being convinced that p . And what goes for conviction – or belief – goes equally for other propositional attitudes. Feeling surprised that p is a way of being surprised that p , feeling angry that p is a way of being angry that p , and so on. When you feel surprised, what you feel is just that – namely, surprise. When you feel angry, what you feel is anger. Similarly, when you feel conviction, what you feel is conviction – that is, belief in the binary sense.

5. An Argument from Normativity

Suppose we acknowledge the distinction between occurrent and dispositional concepts of belief. The next objection I want to consider is that neither occurrent feelings of conviction nor dispositions to feel conviction are sufficient for belief unless they play the right causal role in the subject's psychology. According to Eric Schwitzgebel's (2002) *cluster theory*, for example, the disposition to feel conviction that p is not sufficient for believing that p unless it is part of a larger cluster of dispositions, including dispositions to act and react in appropriate ways. On this view, the madman's feelings of conviction don't have the right dispositional or functional profile to be properly categorized as beliefs.

My complaint is that the cluster theory fails to explain why the madman is irrational. To explain this, we must acknowledge that his feelings of conviction are subject to norms of rationality – and, more specifically, to the same doxastic norms of rationality that govern our beliefs. Moreover, we cannot plausibly deny that the madman's feelings of conviction are beliefs while also maintaining that they are subject to the rational norms that govern all beliefs as such. Hence, we cannot adequately explain why the madman is irrational without classifying his feelings of conviction as beliefs.

To illustrate the point, suppose Karl is a madman who feels convinced that his spouse has been abducted and replaced by an imposter. Given Karl's evidence, this hypothesis is extremely improbable: he has little evidence for it and plenty of evidence against it. Nevertheless, Karl feels convinced that this hypothesis is true. Strangely enough, however, this feeling of conviction is functionally isolated from the rest of his psychology. For example, it doesn't cause him to feel anxiety about his spouse or motivate him to call the police.

In this example, Karl is suffering from an idealized version of Capgras delusion. In real-life cases of Capgras delusion, some of the normal functions of belief may be preserved. Indeed, functionalists sometimes defend the doxastic conception of delusion on the grounds that delusions satisfy minimal functional constraints on belief (Bortolotti 2010). It is questionable whether real-life delusions play a causal role that is similar enough to paradigm beliefs to qualify as beliefs by functional criteria alone (Schwitzgebel 2012). For current purposes, however, we can circumvent this dispute by assuming that Karl's feeling of conviction is disposed to play none

of the causal roles that we associate with the concept of belief. In other words, we can stipulate that Karl's feeling of conviction is "mad" in Lewis's technical sense.

My opponent doesn't deny that this scenario is conceivable. Instead, she denies that the scenario we're conceiving is one in which Karl is aptly described as *believing* the delusional hypothesis. We're assuming that Karl feels convinced that it's true. The phenomenal character of his experience is just like yours would be in the unlikely event that you feel convinced that your spouse has been abducted and replaced by an imposter. Moreover, this phenomenal similarity is sufficient to determine an intentional similarity. It's not just that Karl feels how you would feel if you were convinced the delusional hypothesis is true. We can say more: *Karl feels convinced that the delusional hypothesis is true*. My opponent grants all this. Nevertheless, she denies that this feeling of conviction is aptly described as a belief because it doesn't play the right functional role.

The problem is that my opponent cannot explain what is evidently true – namely, that Karl is irrational. It's epistemically irrational to feel convinced that something is true in the absence of sufficient evidence. And it's practically irrational to be wholly unmotivated to act on what you feel convinced is true. Karl is irrational in all these ways. To explain this, we must recognize that his feelings of conviction are subject to norms of rationality of the same kind that govern our beliefs. Just as it's epistemically irrational to *believe* something in the absence of sufficient evidence, so it's epistemically irrational to *feel conviction* in the absence of sufficient evidence. Likewise, just as it's practically irrational to be unmotivated to act on your *beliefs*, so it's practical irrational to be unmotivated to act on your feelings of conviction. And so on. The norms that govern Karl's feelings of conviction are none other than the doxastic norms that govern our beliefs. Once we recognize this, we cannot plausibly deny that Karl's feelings of conviction are beliefs, since belief is the kind of state that is governed by these doxastic norms.

This is just one instance of a more general argument that feelings of conviction – or dispositions to have them – are sufficient for belief. Here is the argument in outline:

- (1) All feelings of conviction – and dispositions to have them – are subject to doxastic norms of rationality of the same kind that govern our beliefs.
- (2) All mental states that are subject to these doxastic norms of rationality are beliefs.

(3) Therefore, all feelings of conviction – and dispositions to have them – are beliefs.

Anyone who denies that the madman's feelings of conviction are beliefs thereby faces an uncomfortable choice between rejecting one or other of these two premises.

The first premise seems hard to deny. After all, we routinely evaluate our cognitive feelings of confidence, doubt, and conviction as rational or irrational, reasonable or unreasonable, justified or unjustified. It is irrational to feel more or less conviction than is justified by your evidence. Moreover, we cannot make good sense of these evaluations without supposing that feelings of conviction are subject to the same doxastic norms of rationality that govern our beliefs. Other mental states – such as imagination, acceptance, and desire – are subject to their own distinctive norms of rationality. But we cannot adequately explain why Karl is irrational by supposing that his feelings of conviction violate any of these non-doxastic norms of rationality. After all, rationality doesn't require that non-doxastic states – such as imagination, acceptance, and desire – are based on sufficient evidence.

It's worth noting that this poses a serious challenge for non-doxastic theories of delusion, according to which delusions are not beliefs, but imaginings (Currie and Ravenscroft 2002), acceptances (Dub 2017), or other non-doxastic states. After all, imaginings, acceptances, and other non-doxastic states are not subject to the same norms of epistemic rationality as belief. It is not inherently irrational to imagine or to accept a hypothesis that is not adequately supported your evidence. Hence, non-doxastic theories cannot explain the epistemic irrationality involved in feeling convinced by a delusional hypothesis in the absence of adequate evidence. To avoid this objection, non-doxastic theories must deny that delusional subjects feel any sense of conviction that the delusional hypothesis is true; instead, perhaps they are mistaken about their own experience and falsely believe that they do. While this cannot be ruled out in principle, it is a speculative hypothesis that needs substantial empirical support: it cannot be motivated solely by functionalist ideology in the philosophy of mind.

In any case, we can bypass these empirical issues here. After all, it is a stipulation of our example that Karl feels conviction that the delusional hypothesis is true. And it is irrational to feel conviction that something is true in the absence of sufficient evidence. But it is not irrational to

imagine or accept a hypothesis in the absence of sufficient evidence. So we cannot adequately explain Karl's irrationality by supposing that his feelings of conviction are subject to the kinds of norms that apply to these non-doxastic states.

Another option is to reject the second premise by denying that only beliefs are subject to doxastic norms of rationality. It may be said that *judgments* are subject to the same doxastic norms of rationality as *beliefs*. As we saw in §4, however, this distinction between belief and judgment is spurious. We can distinguish between occurrent and dispositional beliefs, but these are two distinct determinates of a common determinable. Denying this merely obscures the fact that these states are governed by the same kinds of doxastic norms.

As we saw in §2, this is precisely what generates the problem of chauvinism. The problem arises when we restrict the category of belief in ways that obscure normative similarities. We can certainly divide beliefs into subcategories depending on their subject matter, physical realization, or functional role. But it is chauvinistic to deny that a state is a belief when it is governed by the same normative standards that apply to belief in paradigm cases. It is chauvinistic to deny that Martians have beliefs merely because they diverge from ours in physical realization, since their states are governed by the same norms of rationality as our beliefs. For the same reason, it is chauvinistic to deny that madman have beliefs merely because they diverge from ours in functional role. This is to obscure what these states have in common with our own beliefs: namely, that they are governed by the same normative standards of rationality.

6. Belief and Implicit Bias

Eric Schwitzgebel (2010, 2021) denies that feelings of conviction are sufficient for belief. He argues that my view generates the wrong results in cases of implicit bias when someone's actions and reactions reveal that their feelings of conviction don't reflect what they really believe. More specifically, he argues on pragmatic grounds that we shouldn't attribute belief in such cases because doing so lets us off the hook too easily for our implicit biases.

Let's begin with Schwitzgebel's (2010: 532) example of Juliet, the implicit racist. Juliet has studied the literature on racial differences in intelligence and she finds the case for racial equality convincing. She sincerely asserts that all races are equal in intelligence, and she is prepared to

argue the point in conversation. Nevertheless, she displays systematically racist tendencies in her unreflective actions, reactions, and judgments about particular cases. For instance, she tends to regard her white students as smarter than her black students, to be surprised when black students outperform white students, and to assign them lower grades for equal work.

Schwitzgebel presents this as a counterexample to the thesis that feelings of conviction are sufficient for belief. After all, Juliet feels convinced by the egalitarian hypothesis that all races are equal in intelligence. But does she really believe it? According to Schwitzgebel, there is no determinate fact of the matter: it's an "in-between" case. Moreover, his cluster theory is designed to explain this verdict: it's indeterminate whether Juliet has enough of the dispositions in the cluster required to count as determinately believing the egalitarian hypothesis.

Schwitzgebel doesn't claim that Juliet believes the *inegalitarian* hypothesis that there are racial differences in intelligence. On his view, this is another in-between case, since Juliet has some but not all of the dispositions associated with inegalitarian belief. In any case, our question is not whether the disposition to feel conviction is *necessary* for belief. For current purposes, we can treat this as common ground in the debate between Schwitzgebel and me. The question at issue is whether the disposition to feel conviction is *sufficient* for belief.

I remain unpersuaded by Schwitzgebel's alleged counterexample. I find it plausible that insofar as Juliet *feels* convinced by the egalitarian hypothesis, she *is* convinced, since what she feels is *conviction*. Instead of trading intuitions, however, I'll sketch three arguments for my view before criticizing Schwitzgebel's arguments for the opposing view.

My first argument is that Juliet *sincerely asserts* that all races are equal in intelligence. Moreover, she is not confused about what this sentence means. On the contrary, she means exactly what she says. And when speakers mean what they say, their sincere assertions express what they believe. Hence, Juliet believes that all races are equal in intelligence. This link between sincere assertion and belief is articulated in Saul Kripke's *disquotational principle*: "If a normal English speaker, on reflection, sincerely assents to '*p*', then he believes that *p*" (1979: 249).

My second argument is that Juliet *knows* that all races are equal in intelligence. When she affirms this point in conversation and defends it in debate, she expresses what she knows to be

true based on adequate evidence. In general, however, knowledge requires belief: you cannot know what you don't believe. Hence, Juliet believes that all races are equal in intelligence.

My third argument is that Juliet's feeling of conviction is an *epistemically rational* response to her evidence. Her evidence strongly supports the egalitarian hypothesis that all races are equal in intelligence. And the epistemically rational response to such evidence is to *believe* that all races are equal in intelligence. Given that Juliet is epistemically rational in responding to this evidence by feeling conviction that all races are equal in intelligence, we may conclude that this constitutes her *believing* that all races are equal in intelligence. Admittedly, some of her beliefs about particular cases – like her unreflective actions and reactions – are unresponsive to this evidence. Hence, she is not epistemically rational in all respects. Nevertheless, she does display epistemic rationality insofar as she feels convinced that the egalitarian hypothesis is true. And we cannot adequately explain this without supposing that her feeling of conviction reflects what she believes.

Schwitzgebel (2010: 548, n. 1 & 550, n. 9) considers the first two arguments but not the third, although we can anticipate his likely response. His general strategy is to deny that belief is a necessary condition for sincere assertion, knowledge, or epistemic rationality. The suggestion is that some of the theoretical roles that are traditionally assigned to belief can be occupied instead by judgments or feelings of conviction. So, for example, Juliet's assertions are sincere insofar as they express her judgments; moreover, these judgments constitute knowledge; and they are epistemically rational responses to her evidence.

The problem with this strategy is that it threatens to demote the concept of belief from its position of central importance by reassigning its theoretical roles to judgment instead. Moreover, as I argued in §4, it is spurious to deny that judgments are beliefs when they are subject to the same doxastic norms as other beliefs. For both reasons, we should regard judgment as a determinate kind of belief. Not only does this respect the normative overlap between occurrent and dispositional belief, but it restores the concept of belief to its position of central theoretical importance.

Let me now address Schwitzgebel's arguments for the opposing view that Juliet doesn't determinately believe the egalitarian hypothesis. One such argument appeals to our ordinary

practice of belief-attribution – more specifically, our tendency to withhold belief attribution in cases like Juliet’s. As Schwitzgebel notes, “It might be appropriate to criticize Juliet by saying that she doesn’t *really* believe in the intellectual equality of the races or that she doesn’t believe it ‘deep down’” (2010: 539).

In reply, however, we can explain away our tendency to withhold belief attribution in such cases. After all, Juliet doesn’t believe the egalitarian hypothesis *wholeheartedly* or *without conflict*. She also believes – or she acts in some respects *as if* she believes – the inegalitarian hypothesis. Perhaps, as Schwitzgebel suggests at one point, she experiences “fleeting explicitly racist thoughts which she quickly repudiates” (2010: 541). If so, then she has inconsistent beliefs, since she feels convinced by the egalitarian hypothesis in some conditions and the inegalitarian hypothesis in others. In any case, there is some rational tension within her belief system because as she has beliefs about particular cases that conflict with her general belief in the intellectual equality of the races. This is enough to explain our tendency to deny that she unequivocally or wholeheartedly believes the egalitarian hypothesis.

Moreover, belief attribution exhibits a phenomenon known as ‘*neg-raising*’: in saying that someone doesn’t believe that p , we normally communicate that she believes that not- p (Hawthorne et al. 2015). By contrast, when we say that someone doesn’t know that p , we don’t communicate that she knows that not- p . This explains the tendency to deny that Juliet believes the egalitarian hypothesis. In saying that she doesn’t believe it, what we say is literally false, but it communicates something true – namely, that she believes the inegalitarian hypothesis, or that she has other rationally conflicting beliefs about particular cases.

Schwitzgebel also gives a *pragmatic argument* that we give too much credit to implicit racists like Juliet when we say that they have egalitarian beliefs. This, he claims, “risks flattering us and encouraging us toward noxiously comfortable self-assessments” (2021: 363). Schwitzgebel’s concern makes sense given a functionalist conception of belief as a state that conforms approximately to doxastic norms of rationality. On this view, Juliet has the egalitarian belief only if she is disposed to act and react in ways that are minimally rational given this belief. To say that she has the egalitarian belief is thus to give her a positive evaluation. Since she violates

many of the rational requirements that apply to egalitarian beliefs, however, it is questionable whether she merits any such positive evaluation.

In reply, however, this objection has no force against the phenomenal conception of belief. On this view, belief is a state that is subject to doxastic norms of rationality in virtue of the way it disposes you to feel. There are no rationality constraints on belief, since there is no guarantee that your beliefs conform even approximately to the doxastic norms that govern them. As the madman example shows, believers can be as bizarrely irrational as you like. To say that someone believes something is not thereby to make any positive evaluation of their rationality. On the contrary, it is merely to say that they are subject to evaluation by standards of doxastic rationality that apply to all beliefs as such.

To say that Juliet has an egalitarian belief is not to imply that it functions in all or any of the ways that it rationally should. It is merely to say that doxastic norms of rationality are the appropriate standards to use in evaluating whether her egalitarian belief functions as it should. Indeed, when we consult these norms, we find that a mixed evaluation is appropriate. Juliet's egalitarian belief is rational in some respects: for instance, it is an epistemically rational response to her evidence. At the same time, however, Juliet's egalitarian belief is not rational in all respects, since it fails to exert an appropriate rational influence on her unreflective actions, reactions, and beliefs about particular cases.

The result is mixed: Juliet satisfies some of the rational requirements that govern her egalitarian belief, while violating others. So it is not true – as Schwitzgebel (2021: 361) alleges – that attributing the egalitarian belief to Juliet lets her off the hook for her implicit racism. On the contrary, it enables us to articulate her rational failings more precisely. We gain no insight into her situation by denying that she holds the egalitarian belief or by denying that there is any determinate fact of the matter. Instead, this obscures the issue by shielding her from evaluation with reference to doxastic norms of rationality that apply to all beliefs as such. Instead of denying that Juliet has egalitarian beliefs, we should recognize that her egalitarian beliefs are *dysfunctional* in the sense that they don't play all the causal roles that they rationally should play.

In sum, we do not let Juliet off the hook for her implicit racism by attributing egalitarian beliefs to her. On the contrary, we hold her accountable for violating the normative standards of

rationality that apply to all beliefs as such. In denying that Juliet has egalitarian beliefs, Schwitzgebel deprives us of the resources for holding her accountable in this way. So, ironically, it is his own view that lets Juliet off the hook. Of course, he can criticize her beliefs, actions, and reactions about particular cases on the grounds that they are driven by racist prejudice, rather than evidence. But he cannot adequately explain why Juliet merits criticism of a different kind from an explicit racist who unequivocally believes in his own racial superiority. Juliet knows better than this, although her egalitarian knowledge fails to exert a rational causal role in regulating her beliefs, actions, and reactions about particular cases.

The problem emerges even more starkly in the mirror image case of an explicit racist, Julian, who feels convinced of his own racial superiority, but keeps his racial prejudice hidden: his intellectual feelings are compartmentalized in such a way that he never engages in differential treatment of members of other races. On Schwitzgebel's cluster theory, it is at best *indeterminate* whether Julian believes the inegalitarian hypothesis, since he lacks many of the dispositions in the cluster that we associate with belief. But this means we cannot criticize Julian for holding racist beliefs violate norms of epistemic rationality: they are driven by racial prejudice, rather than objective evidence. Of course, Schwitzgebel might say that Julian is criticisable for making *judgments* that violate epistemic norms of rationality. Once again, however, it's spurious to deny that these judgments are beliefs when they subject to all the same rational norms. Instead, we should recognize that Julian's feelings of conviction are sufficient for belief.

Schwitzgebel's cluster theory groups Julian and Juliet together as in-between cases that display some but not all of the dispositions associated with egalitarian and inegalitarian belief. Both are rationally criticisable insofar as their dispositions are splintered and not uniformly aligned with evidence. Pretheoretically, however, there is an important normative difference between these cases. Juliet is epistemically rational in believing the egalitarian hypothesis, although she is criticisable because her egalitarian belief fails to regulate her unreflective actions, reactions, and beliefs about particular cases. In contrast, Julian is epistemically irrational in disbelieving the egalitarian hypothesis, although he acts in all respects as if he believes it. The phenomenal conception of belief allows us to draw these fine-grained distinctions between these cases, which the cluster theory lumps together.

7. Methodological Reflections

What is at stake in the disagreement about whether Juliet has egalitarian beliefs? Schwitzgebel and I agree that she feels convinced that all races are equal in intelligence, and that she acts accordingly in her more reflective moments, although she displays racist tendencies in her spontaneous actions and feelings about particular cases. It is only when we describe Juliet's situation using the word 'belief' that any disagreement emerges. So why isn't this disagreement a merely verbal one?

Both Schwitzgebel and I agree that we should use the term 'belief' in a way that marks out an important theoretical category. After all, the concept of belief plays important theoretical roles in philosophy as well as ordinary discourse. Nevertheless, we disagree about which are the most important theoretical roles for the concept of belief.

According to Schwitzgebel, the most important category in the vicinity of our ordinary concept of belief is a multi-track disposition that includes not only intellectual dispositions to feel conviction, but also reflective and unreflective dispositions towards action and reaction. For this reason, he suggests, we should endorse the cluster theory, according to which belief is the state of having enough of the dispositions in the overall cluster that we associate with belief.

In response, Tamar Gendler (2008) has offered a theory of belief that privileges reflective dispositions over unreflective ones. She argues that we should reserve the word 'belief' for states that guide reflective behavior in ways that are sensitive to evidence, while coining the word 'alief' for states that guide unreflective patterns of behavior in ways that are more habitual and insensitive to evidence. Thus, she writes, "beliefs change in response to changes in evidence; aliefs change in response to changes in habit" (2008: 566). As Schwitzgebel (2010: 539) notes, however, this distinction seems overdrawn: our beliefs are not always responsive to evidence, while our habits are sometimes responsive to evidence. So there is no clear and principled basis here for privileging reflective behavior in a theory of belief.

My own view is rather different. I don't accept Gendler's claim that our beliefs must be responsive to evidence, since I deny that there are any rationality constraints on belief. On my view, all beliefs are governed by norms of doxastic rationality, but there is no guarantee that

beliefs must always conform to these rational norms even in some approximate or minimal way. To say that someone believes something is just to say that they *should* conform to these rational norms regardless of whether they actually do or potentially could. Our tendency to comply with these rational norms depends contingently on how rational we are. We humans are imperfectly rational creatures: we are neither perfectly rational nor perfectly irrational. As such, we can predict and explain each other's behavior more accurately by assuming that we are imperfectly rational in our beliefs, desires, and actions. But this doesn't mean it's an analytic or conceptual truth that any agent with beliefs and desires is at least imperfectly rational. At best, this is a synthetic truth about human believers that has some predictive and explanatory power.

My view is not that our beliefs manifest themselves only in reflective actions and reactions, rather than unreflective ones. On the contrary, I insist that they *should* manifest themselves in both ways. My concern is to preserve the resources for criticizing implicit racists whose egalitarian beliefs don't play the rational causal roles they should play in regulating unreflective actions, reactions, and beliefs about particular cases. More specifically, their egalitarian beliefs violate the doxastic norms of rationality that apply to all beliefs as such.

This enables me, unlike Gendler, to avoid the charge of marginalizing the importance of unreflective actions and reactions. It is morally significant that Juliet displays racist patterns of action and reaction that conflict with her general belief in the intellectual equality of the races. And yet none of this undermines the claim that she has this egalitarian belief in the first place. Indeed, this enables us to criticize Juliet on the grounds that her egalitarian belief doesn't play all the functional roles in her psychology that it rationally should play. It is rationally responsive to evidence, but it doesn't play a sufficiently rational role in regulating her unreflective actions, reactions, and beliefs about particular cases. So attributing the egalitarian belief does not let Juliet off the hook for her implicit biases. On the contrary, it explains what is wrong with them.

The whole point of saying that Juliet believes the egalitarian hypothesis is to settle which normative standards of rationality are appropriate in evaluating her attitude. If her attitude towards the egalitarian hypothesis is not belief, but some non-doxastic state of acceptance, imagination, or desire, then doxastic standards of rationality are the wrong standards to apply. And if it is indeterminate which attitude she has, then it is indeterminate which standards we

should apply. All of this seems wrongheaded to me. Juliet is criticisable because her egalitarian belief doesn't play the rational functional role it should play in regulating her unreflective beliefs, actions, and reactions in particular cases. Rather than encouraging a comfortable view of ourselves, attributing the egalitarian belief allows us to articulate precisely which normative criticisms are appropriate in cases of implicit bias.

References

- Block, N. 1978. Troubles with Functionalism. *Minnesota Studies in the Philosophy of Science* 9: 261–325.
- Bortolotti, L. 2010. *Delusions and Other Irrational Beliefs*. Oxford University Press.
- Chalmers, D. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.
- Chalmers, D. 2011. Verbal Disputes. *Philosophical Review* 120: 515-566.
- Crane, T. 2013. Unconscious Belief and Conscious Thought. In *Phenomenal Intentionality*, edited by U. Kriegel, 156–173. Oxford University Press.
- Currie, G. & I. Ravenscroft. 2002. *Recreative Minds: Imagination in Philosophy and Psychology*. Oxford University Press.
- Dub, R. 2017. Delusions, Acceptances, and Cognitive Feelings. *Philosophy and Phenomenological Research* 94: 27–60.
- Fodor, J. 1987. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press.
- Gendler, T. 2008. Alief in Action (and Reaction). *Mind and Language* 23: 552–585.
- Harman, G. 1987. (Nonsolipsistic) Conceptual Role Semantics. In *New Directions in Semantics*, edited by E. LePore, 55–81. Academic Press.
- Hawthorne, J., D. Rothschild, & L. Specter. 2015. Belief is Weak. *Philosophical Studies* 173: 1393–1404.
- Helton, G. 2020. If You Can't Change What You Believe, You Don't Believe it. *Noûs* 54: 501–26.
- Kaplan, M. 1996. *Decision Theory as Philosophy*. Cambridge University Press.
- Kripke, S. 1979. A Puzzle About Belief. In *Meaning and Use*, edited by A. Margalit, 239–83. Reidel.
- Lewis, D. 1983. *Philosophical Papers, Vol. 1*. Oxford: Oxford University Press.
- Lewis, D. 1999. *Papers in Metaphysics and Epistemology*. Cambridge University Press.
- Peacocke, C. 2007. Mental Action and Self-Awareness (I). In *Contemporary Debates in the Philosophy of Mind*, edited by B. McLaughlin & J. Cohen, 358–76. Blackwell.
- Schwitzgebel, E. 2002. A Phenomenal, Dispositional Account of Belief. *Noûs* 36: 249–275.

- Schwitzgebel, E. 2010. Acting Contrary to Our Professed Beliefs or the Gulf Between Occurrent Judgment and Dispositional Belief. *Pacific Philosophical Quarterly* 91: 531–553.
- Schwitzgebel, E. 2012. Mad Belief? *Neuroethics* 5: 13–17.
- Schwitzgebel, E. 2021. The Pragmatic Metaphysics of Belief. In *The Fragmented Mind*, edited by D. Kindermann & C. Borgoni, 350–375. Oxford University Press.
- Smithies, D. 2012a. The Mental Lives of Zombies. *Philosophical Perspectives* 26: 343–372.
- Smithies, D. 2012b. The Normative Role of Knowledge. *Noûs* 46: 265–288.
- Smithies, D. 2013a. The Nature of Cognitive Phenomenology. *Philosophy Compass* 8: 744–754.
- Smithies, D. 2013b. The Significance of Cognitive Phenomenology. *Philosophy Compass* 8: 731–743.
- Smithies, D. 2019. *The Epistemic Role of Consciousness*. Oxford University Press.
- Smithies, D., P. Lennon, & R. Samuels. 2022. Delusions and Madmen: Against Rationality Constraints on Belief. *Synthese* 2000, 223: 1-30.
- Stalnaker, R. 1984. *Inquiry*. Cambridge University Press.
- Stich, S. 1978. Beliefs and Subdoxastic States. *Philosophy of Science* 45: 499–518.
- Velleman, D. 2000. On the Aim of Belief. In *The Possibility of Practical Reason*, edited by D. Velleman, 244–81. Oxford University Press.
- Wedgwood, Ralph. 2007. Normativism Defended. In *Contemporary Debates in the Philosophy of Mind*, edited by B. McLaughlin and J. Cohen. Blackwell.
- Williamson, T. 2000. *Knowledge and Its Limits*. Oxford: Oxford University Press.