

11

Risking Belief

John Schwenkler

1 Introduction

It seems safe to assume that you're reading this chapter because you think it a good idea to do so. Less safe, perhaps, is the assumption that you expect to learn something from it—though I'll flatter myself in thinking that you regard this as a real possibility. There are at least different forms that this learning could take. Maybe my arguments will lead you to consider a question that you'd never before had an opinion on, and you'll come to know something about it. In that case you'll gain something valuable, and the only thing you'll pay for it is your time and attention—unless, perhaps, your mind is currently at capacity and the new knowledge pushes something else out of the way. Somewhat less likely, at least if my own past is any guide, is the possibility that I will convince you that one of your current opinions is mistaken, and you'll come to believe (or have greater confidence in) the opposite, or at least no longer believe (with the same confidence) the thing you once did. In that case the bargain is even better: you put in your time, get back some insight, and get some garbage taken out too. Either way, it seems rational to go ahead and trust that learner's instinct: you'll be better off at the end.

Or will you? *Sometimes* doing philosophy has the effects I just described. The things we read are rigorous and insightful, and they prompt us to challenge our preconceived opinions and come to a deeper understanding of the matters under consideration. Other times, however, the effects are less beneficial. When a philosopher's conclusions are mistaken, or her arguments invalid, then if we are moved in the direction of her position it may be by means of mere persuasion, rather than learning, that this happens, and the process may result in our understanding things *less* well than we did beforehand. (There are, of course, also the many times when doing philosophy has no effect on our beliefs at all.) How can you anticipate which way the present experience is going to turn out? And how should this anticipatory judgment affect your assessment of whether or not to keep on reading?

Here is an answer that you, as a professed lover of wisdom, likely think you simply *cannot* give to those questions: that if you were to anticipate that my conclusions are false, this *would* give you a prima facie reason to stop reading this chapter and take up some other activity instead, so that my arguments didn't impoverish your epistemic situation. You, a philosopher, will likely regard this position as counseling a decidedly unphilosophical form of dogmatic closed-mindedness—and if there is

anything that marks a philosopher, it is the way she is open to correction, engages all comers, and follows the argument where it leads (Kelly 2011). How can you do any of this, if you close yourself off to an argument just because you don't agree with where it is going? But then—wouldn't you have a reason *not* to keep reading this chapter, if that is what it is going to convince you is sometimes a reasonable thing to do? And this is, in fact, exactly what I mean to attempt.

2 The Puzzle of Doxastic Transformation

Let's take a step back. Define a *conversion* as an event that satisfies the following criteria:

- It brings about a *change in some attitude* or attitudes¹ of a person.
- The changed attitude is, or essentially involves, *commitment* to the world's being a certain way (where one's own life is part of "the world" in the relevant sense). The principal contrast I have in mind is with the non-committal states of credence, preference, and desire: thus belief and intention are paradigmatic commitments on my way of thinking, though in this chapter I'll consider only the case of ("theoretical"²) belief. Related puzzles about the revisability of intention have been treated extensively elsewhere in the literature.³
- The changed attitude is *central to the person's worldview or way of life*.
- There is a considerable *distance between* the new attitude and the old one.

One of the things that makes conversion philosophically interesting is the difficulty in seeing how an event that satisfies the above criteria could consist in anything more than a *brute*, or extra-rational, change in the attitude or attitudes in question.⁴ As Bas van Fraassen has observed, the questions that arise here are similar to classic questions about the rationality of scientific revolution: if attitude (or theory) *Y* appears simply *ludicrous* by the lights of attitude (or theory) *X*,⁵ and *X* explains well enough most of the phenomena that *Y* is supposed to account for, then it can be difficult to see what *rational* means there could be of getting a person committed to *X* to abandon this commitment and commit herself to *Y* instead. While I think these questions are very important, my focus in this chapter will be somewhat different. I will take it for granted that it *is* possible for conversion to take place through the proper operation of one's rational faculties rather than by means of brute force or some other non-rational process. And the question that I am going to raise

¹ For the sake of simplicity, I'll mostly speak in what follows of a singular "attitude" that is the locus of conversion.

² See Marušić and Schwenkler (2018) for the distinction between practical and theoretical belief.

³ For a start, see the essays collected in Bratman (2018).

⁴ This is roughly the problem that John McDowell raises in his important paper "Can there be external reasons?," which has had a significant impact on my thinking about these matters. McDowell introduces the term *conversion* as a placeholder for "the idea of an intelligible shift in motivational orientation that is exactly *not* effected by inducing a person to discover, by practical reasoning controlled by existing motivations, some internal reasons that he did not previously realize he had" (McDowell 2001: 102).

⁵ If, that is, "the new view is literally absurd, incoherent, inconsistent, obviously false, or worse—meaningless, unintelligible—within the older view" (Van Fraassen 2004: 72).

concerns whether, and in what circumstances, the seeming falsity of a view could reasonably be counted as a reason not to take the *risk* of engaging with arguments in its favor, or with experiences with the potential to convert one to it, since by one's current lights the *result* of having one's mind changed by those arguments or experiences would be, as L. A. Paul puts it, a kind of *cognitive impairment*:

When is it the case that, before I make a decision to become a different sort of self, I can rightly regard my future self as cognitively impaired, relative to my current self? Should I, from the perspective of my current self with her current preferences, regard any dramatic change of my preferences, especially transformative changes to my core personal preferences, as a kind of cognitive impairment? Where is the line between revising one's preferences in response to experience such that one autonomously learns from the experience, versus having one's preferences controlled by the experience? (Paul 2015: 806)

While Paul frames her puzzle in terms of a change in one's personal preferences, a similar worry arises when the attitudes that stand to be changed by an experience are *commitments* in the sense defined above. Suppose, for example, that you believe that God exists, that government should represent the will of the people, or that it is wrong to eat the flesh of animals when meatless options are readily available. And now suppose that I propose to present an argument, or expose you to some sort of experience, that has a good chance of getting you to abandon this belief, and perhaps to believe the opposite thing instead. By your current lights, the transformation I'd thereby bring about would be a cognitive impairment, not an improvement. And while of course it is true that, by the lights of the person you'd be in the wake of this transformation, the change would have improved your cognitive situation quite a lot, by your current lights it seems to you that this is just what you *would* think, as the meat-eating, democracy-hating atheist you would unfortunately have become.⁶ In such a situation it seems reasonable to ask: is that really the sort of thing that you should risk having happen to you? Doesn't the perceived badness of the outcome give you at least *some* reason to keep it from coming about?

The *puzzle of doxastic transformation* that I'm trying to get you to take seriously has a similar structure to the puzzle of personal transformation that Paul puts forward in *Transformative Experience* (Paul 2014). Paul's puzzle is supposed to arise when (1) a person has to make a choice that is based on considerations of her own subjective well-being, but (2) since the choice involves a sort of experience that she's never had before, she's not in a position to weigh its value or determine how she is likely to respond to it. Moreover, (3) the person knows that if the experience does

⁶ Again, van Fraassen puts it well: "how can we tell that what we see as material welfare in that kind of future will be cognized as such then? And if it is not, what about a future in which we are by our present lights well off, and by our lights then miserable or suffering a great loss? Perhaps we can suppose that in contemplating the decision I see myself laughing and outwardly cheerful in that future. But given the opacity to me now of what my words and body language then really mean, I really have no access to how things really seem to me then. Will this cheerfulness be the false face adopted in despair of ever regaining what I have lost? None of these difficulties appear in the ordinary everyday case where we can assume that our future way of thinking about ourselves will be the same as it is now, factual details aside. That, however, is precisely what we cannot assume here. There is therefore no rational way of deciding upon such a transition, if rational means 'rational by the lights we have beforehand'" (2004: 101–2). This last conclusion (following "therefore") is one I want to resist.

alter her preferences, then she'll be satisfied living in a way that would not satisfy her at all presently. And what makes Paul's puzzle puzzling is the intuition that, given the limitations in (2) and the asymmetry in subjective satisfaction that arises due to (3), the person isn't able to choose wisely as per (1).

As I noted just above, an important difference between my puzzle and Paul's is that in the cases that interest me a person's *beliefs* are at stake, and not merely her core preferences (though of course a conversion might alter those too). But by my lights this makes the puzzle of doxastic transformation only more pressing, since beliefs aren't *subjective* in the way that (at least some) preferences are. Because of this, the worry that through some transformative event I may end up *preferring* things wildly different than the ones I currently prefer is not as serious as the worry that I will end up getting things *wrong* by coming to have a lot of false beliefs. My puzzle, then, arises when (1') a person has to make a choice that is based on considerations of her own cognitive well-being, but (2') she's not in a position to say in advance whether the choice will lead to an improvement in her cognitive situation rather than an impairment of it. Moreover, (3') the person knows that if her beliefs *were* to change, whether rationally or not, then she'd regard this change as an improvement. And so the puzzle concerns the fact that, given the limitations in (2') and the asymmetry in cognitive perspective that arises due to (3'), the perspective from which the person must make the choice in (1') appears inadequate to the task. At least as far as epistemic goods are concerned, the rational choice to make is the one most likely to lead to true beliefs (or knowledge or understanding or wisdom, etc.—I take no standing on the hierarchy of epistemic value) by way of a rational process. But the person making the choice is barred from scrutinizing directly the epistemic standing of the potentially transformative process, and her attempt to evaluate which beliefs are true and which belief-forming processes are epistemically upstanding will necessarily reflect her current perspective.

Perhaps you think that this puzzle is not really puzzling at all, or that you have on hand an easy solution to it. In that case I invite you to go ahead, take the risk. Let's see what I can do to change your mind.

3 An Objection: The Whole Thing Is Badly Conceived

"*Choosing* what to believe? This reeks of doxastic voluntarism. And what's with this business about treating the potential truth and falsity of our beliefs as relevant considerations in our decision-making? Has anyone actually *done* this, ever? All this feels like a philosopher's pseudo-problem: irrelevant to everyday life, with its supposed force dependent on a questionable framing and a bunch of unsupported assumptions."

Hold on there. I agree that doxastic voluntarism is a Bad Thing—and if anything in my setup of the puzzle of doxastic transformation presupposed it, then this would indeed be sufficient reason to think the puzzle ill-conceived. But I *haven't* presupposed any such thing. The doxastic voluntarist holds that belief itself is subject to the will—that is, that it's possible to choose, or at least to exercise some degree of

voluntary control over, our doxastic commitments themselves, in something like the way that we exercise voluntary control over our bodily movements. And what makes this impossible is that it's in the nature of belief to "aim at truth," and thus to be responsive only to considerations that bear on the truth or falsity of the matter in question, whereas whether one *wills* or *wishes* to believe something isn't usually such a consideration at all. This doesn't, however, mean that our doxastic situation isn't to a great extent the product of voluntary control, since the choices we make about such things as what to watch or read, what to study and where to do it, whom to be in conversation with and about what topics—all of which concern matters that are clearly under the authority of the will—have an undeniable influence on what we go on to know and believe.⁷ Care about who won the Knicks game last night? Turn on SportsCenter. Interested in international affairs instead? Pick up the *New York Times*. Want to learn more about the forces that govern the interaction of subatomic particles? Maybe you should take a course in physics. None of this implies that you can learn about these things, or influence in any way what you believe about them, simply by some internal act of the will.

The second thing is trickier, since I *do* want the puzzle I am raising to be anchored in questions that concern us in the course of everyday life, and it is fair to complain that the problem I've outlined does not capture very well the "lived" structure of any real-life quandaries. That is, however, partly because this representation of the puzzle *abstracts* from so much of what we take to be relevant in making choices with the potential to influence our future beliefs—considerations like wealth, power, comfort, convenience, the desires of our loved ones, and so on. And this abstraction has a purpose, namely to shift our attention away from those considerations and toward the question that is my focus here, i.e. that of how (if at all) we should regard the value of true belief *itself* as a factor in our practical deliberations. I think it is clear that most of us *do* care, to some degree at least, simply about getting things right, and we are disposed to make choices that help us to do this. Doesn't this generate rational pressure to avoid making choices that threaten to leave us epistemically worse off?

This is not to say that the risk of getting things wrong usually presents itself as such as a salient factor in practical deliberation. But *sometimes* it does. Think, for example, of the secret agent who is tasked with infiltrating an extremist group whose ideology she thoroughly rejects, and who must account for the possibility that the time she spends in close contact with these extremists will soften her opposition to their beliefs. A similar situation may arise for an academic researcher who is interested in studying the effects of propaganda or media bias: isn't it sensible for her to weigh the risk that in exposing herself to a barrage of the messages whose rhetorical force she wishes to investigate, *she'll* be influenced by those messages to some degree, and emerge from her research having "learned" that they contain a good deal of truth? And then there's the case of the dogmatic religious fundamentalist, who worries that her (or her child's) salvific worldview could be harmed by exposure to a too-sympathetic presentation of the supposed discoveries of modern science. *That* is an

⁷ All these are examples of what Pamela Hieronymi (2006) helpfully calls "managerial control" of one's attitudes.

instance of closed-minded anti-rationalism, you are likely inclined to say. And I agree. But is this simply because of the way it involves regarding dramatic change in one's core beliefs as a kind of cognitive impairment?

4 Four Inadequate Responses to the Puzzle

4.1 *Stand Pat*

Let's start by considering the position of our imaginary dogmatist (who is, it should be emphasized, merely a philosopher's caricature of her real-life counterparts). In deciding whether or not to make a choice that promises to change a certain core belief of hers, the dogmatist treats the truth of that belief as a fixed point in her practical reasoning. As long our dogmatist reasons in this way, the prospect of a changed belief is guaranteed to be, not ruled out (for there may be further considerations that speak sufficiently in favor of openness to the possible transformation), but regarded as something that should *so far* be avoided.

As I've just indicated, it doesn't follow from this way of thinking that the dogmatist will *never* choose in favor of potential doxastic transformation. Sometimes she might make such a choice on the basis of non-epistemic considerations, such as the value of wealth, popularity, or power, or of sharing a valuable experience with her friends or family. But it is possible as well for epistemic considerations to be taken up in this sort of calculus: while having a mistaken belief concerning a certain matter might be counted as very bad, this particular badness could be outweighed by the goodness of gaining correct beliefs about a whole lot of other things. The rational space for these choices will shrink, however, as the belief at stake becomes more central to the dogmatist's identity or way of life, since then the epistemic disvalue of losing this belief will appear much greater, and the conception of the world that is informed by this belief will assign less value to goods that are in tension with it. In the limit, the prospect of losing this belief will appear as a kind of personal extinction, such that *nothing* in the world could be worth the risk of giving it up.

Is it ever rational to take the attitude of the dogmatist toward the possibility of doxastic transformation? As I will discuss just below, I am inclined to think that sometimes it could be. It is not plausible, however, that this attitude could be the rational one to take *in general*, as it counsels a form of closed-mindedness that makes it difficult to respond appropriately to opportunities for learning. If my conclusion in this chapter is correct, it is indeed reasonable to be vigilant about the possibility of future cognitive impairment. But it seems quite clear that the situation in which one's mind is changed, even on an important matter that is central to one's identity or way of life, must not *always* be viewed in prospect as something that should so far be avoided.⁸

⁸ Though she appeals to the value of having a stable perspective over time rather than the value of true belief itself, I take Sarah Paul's (2015a; 2015b) defense of the rationality of doxastic self-control to be vulnerable to this objection. While *sometimes* the importance of diachronic continuity might provide sufficient reason to refuse to reconsider a belief in the face of countervailing evidence (or expose it to such evidence when one takes it to lie around the corner), this cannot *always* be the case, as then our dogmatist would be rational in remaining steadfastly closed-minded.

4.2 *Risk It*

The next response we'll consider says that the only rational response to the possibility of doxastic transformation is an unhesitating open-mindedness. According to this position, it is simply a mistake to approach a doxastically transformative choice by worrying about the possibility that it will leave one with a false belief—for *that* is something that one will worry about only if she takes for granted the very belief that the transformation might lead her to question or reject, and to reason in this way would beg the question at issue. By contrast, the mark of epistemic virtue is to open oneself up to challenging arguments and unexpected experiences, confident that the *true* position is going to emerge through a process of radically open-minded inquiry.

As attractive as this position can appear at first glance, on further examination it appears to rest on an unrealistic view of the ability of human learners to respond rationally to potential sources of new information. Following Sarah Paul (2015b), let's use the term *epistemic temptation* to describe the sort of situation in which false or unwarranted beliefs strike us as apparently reasonable. These situations are common enough that we have a special class of verbs to indicate the possibility that we may be in one: the stick in the glass of water *looks* like it is bent, the argument *appears* to be valid, she *sounds* like she is telling the truth, it *seems* like this is the correct conclusion to draw from these data. And while of course we are capable of doing a pretty good job of distinguishing reality from appearance in these domains, that ability is far from foolproof, and it is important to be aware of circumstances or domains of inquiry in which epistemic temptation is especially likely to lead us astray. To the extent that a potentially doxastic transformative event appears to have these characteristics, this counts as a *prima facie* reason to avoid it.

That's not, though, to recommend the line of thought that we attributed to our radical dogmatist: that since *P* is true, and in doing *F* I might be led not to believe it, therefore doing *F* is to this extent not recommended. For we can reject the unequivocal advice always to Risk It without simply taking it for granted that our present beliefs are true, and that *anything* that would lead us to change them is therefore a mere temptation. What stands behind our verdict is the much more modest observation that the chance of giving in to epistemic temptation is something we should take seriously in making choices that have the potential to change what we believe. The challenge is to identify how this can reasonably be done.

4.3 *Proceed With Caution*

The extreme responses we have considered so far—according to which we should always Stand Pat, viewing doxastic transformation always as a cognitive impairment, or always Risk It, viewing doxastic transformation always as an opportunity to improve our epistemic situation—are both inadequate. *Sometimes* a doxastically transformative event amounts to a way of learning about the world. But there are other such events that amount to regrettable cognitive impairment instead. How can we approach occasions of potential doxastic transformation in a way that maximizes the first of these outcomes over the second?

An obvious answer is that the choice for a potential doxastic transformation doesn't have to be all or nothing. In choosing to expose yourself to something—an

argument, a powerful experience, a series of relationships with people whose ideology you think is false—that has the potential to change what you believe, you needn't thereby choose to *allow* your beliefs to be changed in this way, no matter what. For it's possible for you to enter into the potentially transformative process with your guard up, thinking carefully about what you encounter and subjecting to rational scrutiny any inclinations you have to change what you think. Doesn't this give you a way to ensure that any doxastic transformation that you do end up undergoing constitute an improvement, rather than an impairment, in your epistemic situation?

Attractive as it may appear on a first pass, this response assumes a too-optimistic picture of the ability of human reasoners to distinguish the true from the false. First, however good you are at thinking carefully and keeping up your rational guard, there are bound to be *some* situations in which your thought becomes careless and your guard slips—situations in which, despite your best efforts, you nevertheless succumb to epistemic temptation. And second, to the extent that you do keep up your guard, this will often involve evaluating potentially transformative influences against the background of the very beliefs that they threaten to change. Our imaginary dogmatist, for example, can't very well count as having opened her mind to scientific discoveries if every time she encounters a scientific claim, she reasons that since it conflicts with her religious worldview, *therefore* it can't be true. Potentially transformative events won't *teach* a person anything if she approaches them so cautiously that they never have a chance of making a difference to the way she understands the world. The advice to Proceed With Caution is probably on the right track, but on its own it offers no real solution to our puzzle.

4.4 *It Depends*

Our first three responses all take the form of universal policies: they say that one should *always* Stand Pat, Risk It, or Proceed With Caution in the face of a potential doxastic transformation. And in considering these responses, we saw that none of these policies is *always* the right one to adopt. There's not, however, any good reason to require that choices of the sort we are considering will admit of a one-size-fits-all solution. Indeed, it seems on the contrary that the way to approach a doxastically transformative choice will *depend* on the epistemic credentials of the belief that such a transformation would lead one to reject, as well as on the nature of the process that would lead one to do this. This is part of why we think, for example, that a scholar is justified in acting to preserve her beliefs in a way that a conspiracy theorist is not: for the scholar has *good reason* to believe as she does, whereas the conspiracy theorist is a paradigm of irrationality. It is also why we think it more justifiable to close oneself off to forms of doxastic transformation that involve, as Paul puts it in the passage that I quoted in Section 2, having one's beliefs "controlled" or *manipulated* than it is to avoid opportunities for *learning* from novel experiences (which may include such things as study, argumentation, and so on). Can appeal to considerations like these show the way out of the difficulties we have raised?

Let's consider first the possibility that the crucial difference lies in the epistemic status of the belief that a doxastically transformative experience promises to change. The idea here may be that to the extent that a person is *justified* in believing something, she will also be justified in believing that coming to believe the opposite

thing will constitute a cognitive impairment. And this is because a person's justification for believing the thing in question will also be her justification for rejecting the possibility that a change to this belief would put her more in touch with how things really are. The difficulty, however, is that there is a kind of justification that can be possessed even by our imaginary dogmatist or conspiracy theorist, who may be able to appeal to any number of considerations that she takes to support her point of view. And by the same token, many of the beliefs that we seem to be the most justified in refusing to risk the loss of are also beliefs that we seem to have very little justification for: the belief in democracy, for example, or in the dignity of humanity, or in the importance of tolerance and open-mindedness. Cases like that of the justified conspiracy theorist suggests that justification for a belief may not be sufficient grounds for refusing to expose it to counterevidence. Cases of the latter sort suggest that this may not be necessary to make such a refusal rational, either. In each case, the appeal to justification is insufficient to solve our puzzle.

The other possibility we need to consider is that transformative processes can be evaluated according to the *means* by which they bring about doxastic transformation. For example, part of what makes it rational for you to keep on reading this chapter even if you would rather not be brought around to its conclusion—or for our imaginary dogmatist to open her mind to scientific evidence despite the worry about where it will lead her—is that encountering scientific evidence and engaging with philosophical argumentation are ways of *learning* how things are. This makes them different from exposure to pure propaganda or various forms of emotional manipulation, which influence or *control* our beliefs in ways that we regard as epistemically problematic even if the beliefs they bring us to have happen to be true. Given this distinction, can't we evaluate the epistemic credentials of potential doxastic transformations, not by considering just the truth or falsity of the beliefs that they might result in, but also by reference to the means by which they might bring these results about?

I'll argue just below that a version of this response is defensible, but for now we should notice an obvious difficulty with implementing it. The advice we are considering says that a person should evaluate an instance of potential doxastic transformation by considering whether or not it involves a process of learning rather than one of mere control or manipulation—but by what criteria, and according to which standards, is one supposed to decide this? Consider the situations of an atheist trying to decide whether to participate in an emotionally charged religious revival, and a Christian fundamentalist trying to decide whether to keep an open mind as she takes a course in evolutionary biology. In each case the deliberator's *current* perspective might lead her to regard the process by which her beliefs would be changed as largely non-rational or insufficiently driven by evidence: the atheist, because she thinks of religious experience as emotionally charged hallucination; and the fundamentalist, because she thinks of science as ideology and of human reason as too deeply flawed to penetrate the mysteries of creation. Yet things would appear quite differently from the perspective that each of them would have if the transformative processes in question were to unfold. Having been brought to religious belief by a powerful experience, our former atheist will see such experience as the revelation of a transcendent reality. Having been convinced of the credentials of science by an open-

mindful course of study, our former fundamentalist will see the methods of scientific inquiry as an appropriate way to supplement and even reshape our religious convictions. How should these competing perspectives be prioritized? We are faced with another version of the original puzzle.

The difficulty here is not that there is *no* way, from the perspective of a given system of belief, to scrutinize the epistemic credentials of a process with the potential to transform that system. It is rather that in order to do this effectively one cannot simply take for granted the beliefs that this process threatens to transform. At least some of these beliefs will need to be “bracketed,” as we sometimes say, in order to evaluate the credentials of the process from an appropriately neutral perspective. However, to the extent that the beliefs in question really are *central* to one’s way of understanding the world, if they are taken out of play there may be not enough left to bring one’s reason to bear on the crucial question. For there is, as van Fraassen reminds us, no such thing as an epistemology that is altogether independent of our presuppositions about the nature of the world and our situation in it:

There is no way to write a theory of cognition while escaping from our general beliefs about what we and our world are like. We cannot construct a presuppositionless theory, a priori, independent of our current science, theology, metaphysics, or whatever else we have accepted as knowledge. But neither can we construct a theory based in our current knowledge base and still make sense of the idea that we might be . . . capable of correctly attaining, through a conceptual revolution, a true insight radically at odds with that current knowledge base.

(2004: 81–2)

A bit later on I will revisit this last position, attempting to identify a way in which the epistemic credentials of a doxastically transformative process *can* be evaluated rationally and in a non-question-begging manner. First, however, I need to dislodge an assumption that made our original problem seem so intractable.

5 A Better Way Forward

So far I’ve represented the puzzle of doxastic transformation in terms of a situation in which a person’s current beliefs lead her to the conclusion that certain dramatic changes in those beliefs would constitute a kind of cognitive impairment. The puzzle arises from the fact that it seems at once unwise to disregard this possibility altogether and unreasonable simply to take one’s present beliefs for granted in weighing the likely costs and benefits of a potential transformation, or evaluating the epistemic credentials of the processes it would involve. Nor, however, can one approach these questions in a manner that abstains from any potentially controversial claims about what reality is like and what ways we have of coming to know it. But *which* of such claims is it reasonable to rely on?

The arguments in Section 4 were all focused on showing that not all the things a person *believes* to be true are appropriately taken for granted in reasoning about potential doxastic transformation. What I will propose now is that with *knowledge*, the situation is different.

A simple case will illustrate the basic idea. Imagine that you believe that you left the stove on before leaving the house, and so you return home to turn it off. And now

imagine further that the stove wasn't left on after all—you turned it off when you finished cooking your meal, but have altogether forgotten having done this. The stove is not on, but your decision to go home and turn it off still makes a kind of sense: you are heading back home because you believe that the stove is on. You are not, however, going home because the stove actually *is* on—this can't be what explains your going home, since it isn't the case at all.

What more would be required for the stove's being on—the *fact that* it is on, as we say—to be what explains why you go home to turn it off? One thing, obviously, is that the stove would need to *be* on: for an explanation of the form "X because Y" can't get off the ground if Y isn't even the case. Yet the stove's being on also can't explain why you are going home unless that fact is related to your action in the right kind of way. Thus if, for example, you turned the stove off before you left the house and then it was turned back on by someone else, then the stove's being on still won't be what explains why you are going home. Here you are in the lucky position of acting from a belief that happens to correspond to how things are, but still it is merely your belief, and not the fact that things are that way, that explains what you are doing.

This sketch of an argument needs a good deal of filling in, but this isn't the place to provide that.⁹ Instead I am going to take this groundwork for granted and develop what I think is the correct conclusion to draw from it, which is that *knowledge* of a fact is what makes it possible for the fact itself to be what explains, and therefore rationalizes, the things we do on the basis of it in a way that mere belief in a fact does not. And this means that someone who *knows* that something is true can respond to the value of believing this, and the disvalue of believing the opposite, in a way that someone who merely *believes* something to be true cannot.¹⁰ In this respect we may contrast the situation of the academic researcher worried about the effects that a close study of extremist subcultures might have on her understanding of the world, with that of a paranoid skeptic who shuts out the "mainstream media" in order to keep himself from being taken in by globalist propaganda. What makes their situations different is not just that one of them happens to be correct while the other is not—for even a conspiracy theorist will sometimes stumble upon the truth. Rather, I suggest that the crucial difference is that, insofar as the researcher has *knowledge* of the worldview that she wishes to keep her research from undermining, the *facts* that she reasons from can serve as her justification for deciding against a course of action that would put her out of touch with those facts. Our skeptic, by contrast, decides as he does only because he *believes* that the world is a certain way. And the mere fact that one believes something does not provide the same kind of justification for acting in ways that maintain this belief that a known fact can provide for acting in ways that keep one cognitively in touch with it.

⁹ I take the necessary groundwork to have been laid by Hawthorne and Stanley (2008) and Hyman (2015: chs 6–8), among others.

¹⁰ Notice that in each case—for the knower as well as the mere believer—it is only the value of true belief itself that I assume to be at stake. The case for a knowledge-centered solution will get only stronger if the value of knowledge exceeds that of mere true belief—though the ability to respond rationally to that added value might depend on knowledge that one knows, which again is not assumed in my argument here.

Let me guess at what you are thinking. Even if this position is correct as far as it goes, as a proposed solution to the puzzle of doxastic transformation it leaves a lot to be desired. That's because a mere believer, no less than a knower, will very often *take* herself to be justified in reasoning, not merely from the fact of her beliefs, but from the believed facts themselves, to a conclusion that she takes those facts to justify. This sort of thing happens *all the time*: e.g. thinking that my keys are in the kitchen drawer, I open the drawer to get them, and it's only after my search comes up empty that I'm inclined to say that I opened the drawer only because I *thought* the keys were there, and not because they actually were. So, at least in the absence of a reliable way of distinguishing knowledge from mere belief, the solution fails to provide the sort of guidance we were seeking.¹¹

There are several things to say here. First of all, notice that even the advice to act *according to one's beliefs*, or those of one's beliefs that are the most justified or confidently held, would have to face its own version of this objection unless we could rule out the possibility of being mistaken as to what one's own beliefs (or one's most justified or deeply held beliefs) actually are.¹² If matters like these are not infallibly self-known, then even a philosopher who claims that it is always ("subjectively") rational to act according to one's current lights will have to allow for the possibility that a person can be wrong about what it is rational for her to do. Second, notice also that nowhere I have claimed that knowledge *of knowledge* is required for one's choices to be rationalized by a worldly fact. What I have proposed is that a person can rationally choose to (do *F* because *P*) only if she knows that *P* is the case, and that if *P* is something she merely believes and does not know, then the most she can rationally choose is to (do *F* because she believes that *P*). On this account, just as it is possible to think that a choice would be rationalized by the facts when really the choice would be grounded merely in one's beliefs, so *if* it is possible to know something without knowing that one knows this, then it is possible to lack knowledge of what it is rational for one to do. But this doesn't mean that one can't *rationally choose to do one thing or another*—not, anyway, unless we assume that matters of (even "subjective") rationality are infallibly self-knowable, which (as we've just seen) is arguably unachievable on *any* reasonable view of the limits of human self-knowledge.¹³

This leads to a third and final point, which is that an adequate response to the puzzle of doxastic transformation does *not* require providing guidance that shows how to decide, in any given case, whether or not to act in a way that invites the possible transformation of one's core beliefs.¹⁴ As I introduced the puzzle in Section 2, I gave it the form of a "How possible?" question: the puzzle seeks an explanation of how it *could be* rational on some occasions to resist potential conversion on the grounds that it would constitute cognitive impairment, and on others to open oneself up to possible conversion on the grounds that it provides an opportunity for learning. If this is the question we are trying to answer, then our

¹¹ Thanks especially to Kyla Ebels-Duggan for pressing me on this point.

¹² For an argument that *none* of our mental states are "luminous" in this way, see Williamson (2000: ch. 4).

¹³ For further discussion of this last matter, see Hawthorne and Srinivasan (2013).

¹⁴ As Nomy Arpaly put it to me, we should not demand that philosophy provide us with a *manual* that will tell us what to do, or believe, given our beliefs, preferences, and facts about the world as we find it.

question is answered by recognizing the special way that knowledge puts us in a position to act in light of the facts, as then we can understand the goodness of an inference like

(K) P , so I should avoid doing F , as this would bring it about that I believe that not- P

in a way that does not carry over to an inference like

(B) I believe that P , so I should avoid doing F , as this would bring it about that I believe that not- P .

The recognition that *sometimes* a person can reason along the lines of (K), and so choose rationally against potential conversion in a way that a person who reasons merely along the lines of (B) could not, speaks directly to our “How possible?” question, even if it does not answer the further question of how to tell *whether* one is in a position to reason in the first way rather than the second. There is of course a corresponding “How possible?” question, as well as any number of more practical ones, concerning how knowledge and mere belief can be distinguished from the first-person perspective—but these matters fall outside the scope of the present inquiry.

Let’s return now to the responses to the puzzle of doxastic transformation that were discussed in Section 4.4. In that section I argued, first, that mere *justification* for a belief is neither necessary nor sufficient ground for deciding that an experience that would change this belief would amount to a form of cognitive impairment. We are now in a position to see why this is. Justification for a belief, at least of the sort possessed by our hypothetical fundamentalists and conspiracy theorists, is not *sufficient* to rationally reject a doxastically transformative experience, simply because justification of this sort is fallible: it is possible to have this sort of justification without thereby *knowing* the thing that one justifiedly believes. And it will not be *necessary* for such a decision as long as justification is not necessary for knowledge—that is, as long as it is possible to know certain things, e.g. that human beings are ends in themselves, without this knowledge depending on further justification.

The other thing I argued in Section 4.4 was that a person cannot reasonably decide whether a given process of doxastic transformation would center on *learning*, rather than mere influence or control of her beliefs, in a way that simply takes for granted the beliefs that the process promises to change—as an atheist might assume that religious experience is entirely hallucinatory, or a conspiracy theorist might think that scientific consensus is a state-sponsored hoax. In addition, I argued, if a person were to “bracket” or set aside *every* part of her worldview that a process promises to change, then there might not be enough of that worldview remaining for her evaluation of the process to appeal to. The upshot was that we could not solve the puzzle of doxastic transformation simply by saying that one should evaluate the kind of process that a potential doxastic transformation would consist in.

But there was something troubling in this conclusion. For the sort of reasoning whose possibility this argument calls into question seems like a sort of reasoning that people *must* sometimes engage in, and it would be startling to discover that there is no way of engaging in it rationally. Suppose, for example, I have no idea how old the Earth is, and am considering two possible courses of study that would lead me to

different conclusions about its age. If one of these courses of study centers on careful reasoning and the consideration of scientific evidence, while the other takes its bearing from influential myths and centers on patterns of inference that most experts reject, then this *does* seem like a good reason to favor the former course of study over the latter—even as we concede that *if* the conclusion of the second course of study is correct, then mythology will have been underrated, and scientific reasoning over-rated, as a means of obtaining geological knowledge. Does it follow from this concession that we can't appeal to the importance of evidence-based reasoning in choosing to avoid doing things that would lead us to question this importance?

We are now in a position to respond to this challenge. What makes it possible to appeal to the importance of evidence-based reasoning in evaluating the epistemic credentials of a potentially transformative process is our *knowledge* that critical reasoning, and careful consideration of publicly observable evidence, are good ways of learning about things like the age of the Earth, while endorsing a literal reading of ancient myths is not. Since we know these things about what is and is not a good way to learn about the world, we can reason from these very facts to knowledge about the kinds of doxastic transformation that are likely to teach us things (i.e. to give us new knowledge), and not merely manipulate us into believing things that might or might not correspond to the facts. In reasoning about what to do in the face of a potentially transformative process, including about the credentials of that process itself, *the things one knows do not have to be bracketed*.

This last point also allows us to address one more lacuna in the proposal I have put forward in this section. My concern in this chapter has been mainly with understanding the potential rationality of choices *against* potential doxastic transformation—that is, choices *not* to go in for a doxastically transformative experience, on the grounds that such an experience would worsen one's epistemic situation. I have argued that such a choice can be rational only if it is a response to the truth of the belief in question, i.e. to a fact that one knows to be the case and not merely a belief that one holds. But I haven't said anything about what might make it possible to see a doxastically transformative choice as having *positive* value—that is, as positive with respect to the value of the belief it would result in, and not merely the other things that might be taken to recommend it. And it is still hard to see what could account for this: after all, a doxastic transformation is supposed to bring it about that one has a belief that one currently thinks is false, so how is it possible to see such a transformation as potentially good?

I answer that one should welcome a doxastic transformation, seeing it as an occasion for improvement in one's epistemic situation, to the extent that it centers on processes that are ways of learning about how things are. If, for example, you believe but don't know that atheism is true, then on this account you *do* have good reason to engage with philosophical arguments with a good chance of convincing you of the existence of God, since philosophical argumentation is a way of learning how things are. But the same verdict will not hold if you are considering instead whether to start spending time with the members of an emotionally manipulative cult—*not* because the views of that cult are wrong (we are assuming you do not know this), but because their ways of bringing you around to those views would be ways of *controlling* your beliefs, and not a means by which you would *learn* how things are. You might,

210 RISKING BELIEF

of course, still have some reason to spend significant time with the cult members (perhaps they are members of your family, or you are conducting sociological research on this cult's beliefs, etc.), though likely you would do this only in the cautious manner discussed in Section 4.3, and you would *not* be doing it because of the epistemic value of the manipulation that you would be subjected to.

6 Conclusion

Should you have kept on reading this chapter, then? I hope to have shown that it depends on what you knew at the beginning. *If* you had known that the puzzle of doxastic transformation was not a puzzle at all, or that one of the responses to the puzzle that I criticized in Section 4 was an adequate response, then by my own lights the possibility that I would convince you to believe otherwise was something that you could reasonably count in favor of putting the chapter aside. But I don't believe you *did* know that—and, indeed, I think the arguments I have made are sufficient to show that I *know* you didn't know any such thing. If I'm wrong in that, and my arguments have managed to bring you around to my position anyway, then I suppose I owe you an apology—though I do hope the experience was still enjoyable enough, and illuminating with respect to some of the subsidiary claims that I have argued for, to make up for that bit of damage done.

7 Coda

I need to add a confession of disappointment. When I first began working on this project, I was convinced that it *had* to support the conclusion that a person could never be so blinkered by a mistaken worldview that there was simply “no way out” for her—no way, that is, that she *could possibly* choose rationally in favor of an experience that threatened to convert her to something better. My reasons for this conviction were broadly religious: for the possibility of there being “no way out” seems to mean that a person could get to be so badly off that her conversion could happen only against her will—through an event which, *if* she had anticipated what it involved, she *could not* have rationally allowed to happen.

I now think that there is indeed a possibility of this sort. What grounds this possibility is *not* that it is always irrational to make choices that threaten to change our core beliefs, but rather that what it is rational for us to do depends on what we know—and a person with a radically false worldview might be too *ignorant* to reason successfully about which choices will improve her epistemic situation. To the extent that this is true of a person, the only way it can be rational for her to make a choice that would convert her is if there is something else—the invitation of a loved one, perhaps, or the fresh air of a garden, or the promise of getting to Emmaus or Damascus—that grounds her sense of its value, leading her to risk what she has and thereby end up with something much better, something whose worth she could not have recognized in advance.¹⁵

¹⁵ For discussion of these issues, and feedback on earlier versions of this chapter, I am grateful to Nomy Arpaly, Joshua Blanchard, Lara Buchak, Nick Byrd, Nilanjan Das, Josh DiPaolo, Ravit Dotan, Trent

References

- Bratman, Michael. 2018. *Planning, Time, and Self-Governance: Essays in Practical Rationality*. Cambridge, MA: Harvard University Press.
- Hawthorne, John, and Amia Srinivasan. 2013. "Disagreement Without Transparency." In David Christensen and Jennifer Lackey (eds), *The Epistemology of Disagreement: New Essays*. Oxford: Oxford University Press.
- Hawthorne, John, and Jason Stanley. 2008. "Knowledge and Action." *Journal of Philosophy* 105(10): 571–90.
- Hieronymi, Pamela. 2006. "Controlling Attitudes." *Pacific Philosophical Quarterly* 87(1): 45–74.
- Hyman, John. 2015. *Action, Knowledge, and Will*. Oxford: Oxford University Press.
- Kelly, Thomas. 2011. "Following the Argument Where It Leads." *Philosophical Studies* 154(1): 105–24.
- Marušić, Berislav, and John Schwenkler. 2018. "Intending Is Believing: A Defense of Strong Cognitivism." *Analytic Philosophy* 59(3): 309–40.
- McDowell, John. 2001. "Might There Be External Reasons?" In *Mind, Value, and Reality*. Cambridge, MA: Harvard University Press.
- Paul, L. A. 2014. *Transformative Experience*. Oxford: Oxford University Press.
- Paul, L. A. 2015. Replies to Pettigrew, Barnes, and Campbell. *Philosophy and Phenomenological Research* 91(3): 794–813.
- Paul, Sarah K. 2015a. Doxastic self-control. *American Philosophical Quarterly* 52 (2): 145–58.
- Paul, Sarah K. 2015b. 'The courage of conviction'. *Canadian Journal of Philosophy* 45(5–6): 647–69.
- Van Fraassen, Bas. 2004. *The Empirical Stance*. Yale University Press.
- Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford University Press.

Dougherty, Kyla Ebels-Duggan, Anne Jeffrey, John Kvanvig, Enoch Lambert, Sam Lebens, Clayton Littlejohn, Eric Marcus, Laurie Paul, Sarah Paul, Richard Pettigrew, Juan Piñeros, Ted Poston, Jeremy Redmond, Jeff Sebo, and Marshall Thompson. Versions of it were presented at the 2015 Summer Seminar on the Nature and Value of Faith, the 2016 meeting of the Pacific APA, and workshops at Yale University and the University of North Carolina, Chapel Hill.