
This pre-print is the version of the paper first submitted to *Philosophical Studies*, where it's now forthcoming (with some significant modifications).

In Search of Doxastic Involuntarism¹

Abstract:

Doxastic involuntarists, as I categorize them, think it is impossible to form a belief as an intentional action. Considering several ways of elaborating that idea, I argue that none of them makes for an attractive view: if belief-formation is understood in some ways, then involuntarism is false; if in others, involuntarism is insignificant. I also examine several arguments purporting to show that the truth of involuntarism is metaphysically necessary, and I contend that they suffer from the same kind of difficulty: each lacks either soundness or significance. I conclude that involuntarism, if it's to be viable, will need further development of the concept of belief-formation.

Word count: 7,836 words

Introduction

How comforting it would be, and how ennobling, to believe that Mister Rogers is still alive and healthy, and that he often thinks of me fondly, wishing me well. I'd sleep easier at night, with a belief like that. I'd be kinder to others. But it seems I can't form such a belief, despite the practical and moral interests it would serve. Forming that belief isn't an action I can just decide to perform, in willful disregard of all the evidence against it.

That's a dramatic case, since the belief in question would be so flagrantly irrational. Many philosophers, though, think belief-formation is always beyond our volitional control in some important way. Some of them have expressed this idea by claiming that it isn't possible for

¹ I'm sincerely grateful to Bryce Dalbey, Andrew del Rio, Sinan Dogramaci, Daniel Drucker, James Gillard, Steven Gubka, Engel Hawbecker, Sam Krauss, Matthew Matherne, Marissa Neuman, Shmulik Nili, Hannah Trees, and participants in the UT–Austin graduate colloquium, for their discussion of the ideas in this paper.

forming a belief to be an intentional action.² That’s the thesis I’ll be appraising throughout this paper. Call it doxastic involuntarism, or DI.

DI: It is impossible to form a belief as an intentional action.

DI doesn’t deny that our intentional actions can influence the beliefs we form. It doesn’t even deny that particular beliefs could result, sometimes, from actions we perform precisely with the aim of ending up with those beliefs. The belief-formation itself, however, still won’t be an intentional action: only a foreseen consequence of action. So goes the view, anyway.

It has seemed to many doxastic involuntarists that DI is a metaphysically necessary truth, rather than a contingent matter of human psychology. That position, which we could name “metaphysical involuntarism,” is usually supported by arguments that take some essential feature of intentional action, and make it out to be incompatible with some essential feature of believing—its truth-directedness, maybe, or the norms that govern it, or the commitments it involves. Most recently, four such arguments have been given by Nishi Shah and J. David Velleman (2005), Kieran Setiya (2008), Pamela Hieronymi (2009), and Hagit Benbaji (2016).

In the following pages, I criticize those arguments and the involuntarism they underwrite. In Part 1, I present a challenge to DI by considering the three best ways I can think of in which we might understand what it is to “form a belief”: on two of these readings, I argue, DI is false; on the third, it’s trivial. Part 2 inspects the arguments of the metaphysical involuntarists I named above, with similarly unsatisfying results. No available understanding of belief-formation seems capable both of giving it an important place in our cognitive lives and of warranting an involuntarist verdict about it. Though I said I was arguing against involuntarism, then, this paper

² This way of formulating the involuntarist view can be found in Nottelman (2006: 561–57); Setiya (2008: 40); Hieronymi (2009: 157); Steup (2012: 154–58); Benbaji (2016: 1954); and McHugh (2017: 2746–48).

could also be read as simply expressing dissatisfaction with extant forms of the view. If it is to be a viable and interesting position, the category of belief-formation needs serious development.

Before surveying interpretations of DI, let me make one clarification about the topic I'm considering. I'm addressing the question of whether belief-formation could be an *intentional* action. I will leave aside the separate question of whether it could be, more specifically, a *basic* intentional action: an action we perform directly, like raising an arm or winking an eye, without having to perform other actions as parts of it. I need to make this clarification because philosophers interested in belief and the will have often asked the second question, or blurred the first and second together (see Hieronymi 2009: 150–58). Both questions are interesting, but Hieronymi is right to point out that basic actions like wiggling your toes aren't meaningfully more *voluntary* than non-basic ones like reciting a poem (2009: 156). Since I'm on the hunt for a serviceable form of *involuntarism*, I'll follow the lead of the philosophers I'm arguing against: I'll consider the prospect of belief-formation as an intentional action simply speaking, whether basic or non-basic.

Part 1: Candidate Understandings of Belief-Formation

1.1. Acquiring a Belief

How, then, should we understand DI? First of all, it might seem natural to think that what that thesis means by the phrase “form a belief” is simply “bring it about that you have a belief” or, more briefly, “acquire a belief.” That interpretation would result in this revised principle:

DI-1: It is impossible to acquire a belief as an intentional action.

But this first, simple reading won't get us a good form of involuntarism. DI-1 is false: acquiring a belief can be an intentional action.

I'm not saying I can acquire any belief I want. Many beliefs might be quite beyond my powers, like the one about Mister Rogers, above. It's evident from other cases, though, that there's no general impossibility of the kind DI-1 proposes. For instance, we can acquire some beliefs by self-consciously bringing it about that they are *true*. As Richard Feldman (2000: 671–72) points out, if I want to believe that the lights are on in my office, I can get up and flip the light switch. Making a belief true, as a way of acquiring the belief, is a low-tech method, and limited with respect to the content it can get us. It's enough to ensure, though, that intentional belief-acquisition is at least possible.

In our age of ever-accelerating progress, we can imagine more powerful methods. Maybe someday, when pharmaceuticals really come into their own, I'll be able to take a pill and simply find myself a few minutes later with the belief that the lights are on, or that the new cancer treatments I'm undergoing will be effective, or that Mister Rogers likes me. That would be a way of acquiring a belief intentionally even without providing myself with further *evidence* for it (which, you'll notice, is what I'm doing in Feldman's example). In both sorts of cases, though—the low-tech flipping of the light switch, and the medically sophisticated belief pill—I'll be acquiring a belief as an intentional action.

What can doxastic involuntarists say to these seeming counterexamples? They haven't tried to deny the possibility of turning on the lights, or even, necessarily, the possibility of belief pills (see Hieronymi 2009: 150). In fact, for the most part they haven't defended DI-1 at all. What they've done instead is to insist that mere intentional belief-acquisition isn't sufficient for intentional belief-formation. It's uncontroversial, they'll say, that agents can *manage* their beliefs; they can intentionally *acquire* a belief or *bring it about* that they form it (Setiya 2009: 41; Benbaji 2016: 1947; Audi 2015: 27–28; see Alston 1988: 260–84). It's just the

belief-formation itself that can't be intentional. Thus, a distinction between belief-formation and belief-acquisition turns out to be crucial to doxastic involuntarism.

That brings us to the question at the heart of this paper, a question I think should worry involuntarists more than it has: what, after all, *is* belief-formation? What makes it different from just bringing it about that you have a belief? Of course, taking a belief pill *is* a way of intentionally bringing it about that you believe something. But what prevents it from being a way of intentionally forming a belief, as well? Think of it like this: scheduling regular meetings with other musicians to play grunge in your garage is a way of intentionally bringing it about that that you're in a band—but it's also just a way of intentionally forming a band. That is, it's both. There's no conflict between those two ways of understanding the action. Similarly, if some action is a way of intentionally bringing it about that you believe, that doesn't prove it *isn't* a way of intentionally forming a belief.

The involuntarist needs to do more than make the distinction, therefore. She needs to explain it, and to show how the putative counterexamples to involuntarism fall on the wrong side of it. In addition, the distinction she explains needs to be somehow significant: if “belief-formation” ends up looking like a trivial or gerrymandered category, then showing *that* to be involuntary won't establish any interesting view.

That's my challenge to the involuntarist. Responses to it, of one form or another, will absorb my attention in the rest of this essay. In sections 1.2 and 1.3 I imagine and examine some ways of understanding belief-formation, along with corresponding versions of DI; but they fail to meet the straightforward criteria set out in the previous paragraph. In Part 2, I consider some arguments for DI from the metaphysical involuntarists, and I find the same problems manifested in them.

1.2. *Acquiring a Belief by Finding One's Current Evidence Convincing*

We're looking, now, for better ways of understanding belief-formation, and for related versions of DI that don't succumb to the counterexamples seen above. We might try to make progress with a variant of DI engineered specifically to avoid the counterexamples, one like this:

DI-2: It is impossible to form a belief by finding your current evidence convincing, as an intentional action.

DI-2 suggests that the distinguishing feature of belief-formation, the thing missing from mere belief-acquisition, is that it produces a belief by a particular route: by appraisal of the existing evidence. That's a suggestion worth hearing out, and it would helpfully clarify the rough expression of involuntarism in DI. Unfortunately, it won't solve the problems DI-1 had. DI-2 is subject to a revised set of counterexamples.

To be sure, as intended, it's immune from the light switch case. Thought experiments involving belief pills aren't a problem, either: although they aren't cases of acquiring *new* evidence, they also aren't cases of being brought to a belief by evaluating the evidence, which is the process concerned in DI-2. So neither of the earlier cases is a problem. We seem to have plugged the leaks.

There's even a feeling of apriority about DI-2. On the one hand (the involuntarist might say), if someone does find the current evidence convincing, then she'll simply form a belief in the normal way. She'll be moved by the force of epistemic reasons, and that won't count as *intentional* belief-formation. On the other hand, if she doesn't find the current evidence convincing, then she certainly doesn't form a belief *by* finding it convincing. Either way, it seems, she's no counterexample to DI-2. We may be able to manipulate ourselves so that we wind up with certain beliefs. But maybe what we can't do, as Hieronymi says, is "to settle a

question *by* finding convincing reasons that we do not take to settle it” (2009: 184, original italics). That has the ring, again, of an *a priori* truth.

But that feeling of aprioricity is bogus. This much is necessarily true: we can’t form a belief by finding convincing (at time t_1) reasons that we don’t take (at t_1) to settle the theoretical question the belief answers. But it’s only the temporal indexing that makes that principle compelling, together with many similar and unremarkable principles. I can’t read (at t_1) a novel in a language I don’t (at t_1) understand. I can’t shake hands (at t_1) with someone who is standing (at t_1) ten feet away from me. In its barest form, the principle here is just that I can’t φ at t_1 and also not φ at t_1 —and far be it from me to deny that. See how quietly the impossibility vanishes, though, when we allow for different temporal relations between the parts of these sentences. It’s not true *a priori* that I can’t enjoy *Don Quixote* in the original at time t_2 if I have no Spanish at t_1 . There’s no problem shaking hands at t_2 with someone who’s ten feet away at t_1 : it’s really very simple, if you don’t mind taking a few steps in their direction.

In the same way, it’s not at all clear why we couldn’t form a belief by finding convincing (at t_2) reasons that we don’t now (at t_1) take to settle it. But surely it’s *this* question that occupies us when we’re wondering about intentional belief-formation, rather than a question that indexes both verb phrases to t_1 . In the cases that commonly arise, it occurs to some subject that although she doesn’t *now* think there’s good evidence for p , believing p would be a very good thing—it would save lives or win money or improve athletic performance. The question we have is: can the subject then decide to believe, maybe by deciding to find her evidence convincing, and so form a belief in p at a *later* moment? It’s too late for her to believe now, of course, at the moment in which she’s doing her deliberating. To say only that, though, would be to miss the point of the question.

There's no obvious reason, then, to think counterexamples to DI-2 are impossible. And in fact, they're not too hard to find. Take the belief pills considered above. They were ineffective against DI-2 because, while they offered the patient who took them a way of forming a belief, it wasn't a way of forming a belief *by finding the current evidence convincing*. But this limitation can be removed: we only have to consider an expanded line of credal pharmaceuticals. Suppose our enterprising scientists formulate a second generation of belief pills: "evidential standard pills," they could call them. These new pills don't plant the desired belief in the subject directly. Instead, they modify the subject's evidential standards to make the target belief seem well-supported by her evidence. If you want to form a belief in God, for example, you can take a pill that makes the cosmological argument look unanswerable. (It had never really moved you before.) Or, if I want to believe that my football team will win the championship, I can take the pill that makes their losing record seem to support that prediction.

Leaving science fiction behind, consider the advice Pascal (1670/2003: 68) gives his reader after laying out the Wager: "endeavor then to convince yourself, not by increase of proofs of God, but by the abatement of your passions." He counsels such activities as "taking the holy water, having masses said, etc.," promising that these will "naturally make you believe, and deaden your acuteness." He's not telling the reader where to go to get the most convincing new evidence of God's existence. Actually, he explicitly advises against an evidence-gathering approach. Instead, he's recommending a technique for adjusting your evidential standards, finding your current evidence convincing, and thereby acquiring theistic belief in an intentional way. As such, it's also a technique for producing a counterexample to DI-2. It may not be a sure-fire strategy: it's no mean feat, getting your standards to change. But considerations of mere difficulty won't safeguard DI-2: it isn't *impossible* to change your standards. And if Pascal's

reader follows his advice and then begins to believe in God, or if you and I take the evidential standard pills we're prescribed, it seems to me we'll all be doing just what DI-2 says can't be done: intentionally acquiring beliefs by finding our current evidence convincing.

Maybe the involuntarist who likes DI-2 will insist on redescribing the cases. She might say: "When you take the belief pill or the holy water, you're not intentionally *finding* the evidence convincing: you're only intentionally *bringing it about* that you find the evidence convincing. Accordingly, you aren't intentionally acquiring a belief by *finding* the evidence convincing, which is what you'd have to do to be a counterexample to DI-2; you're merely intentionally acquiring a belief by *bringing it about* that you find the evidence convincing."

That response should sound familiar: it's closely parallel to the one which (I said above) involuntarists make to the counterexamples that scuttled DI-1. My response will be the same as it was there: such a distinction can do the necessary work only if the differences it marks can be presented as cognitively significant, and if putative counterexamples to involuntarism can be shown to fall on the wrong side of it. Right now, with DI-2, we're trying to see if this can be done for the distinction between forming a belief and merely acquiring it. As for the distinction that has just been suggested, between finding evidence convincing and bringing it about that one finds it convincing, I don't know how it could be elaborated. If an elaboration is offered, it'll be worth considering. If it's a good one, maybe it can be used to explain how belief-formation differs from belief-acquisition. For now, the distinction remains mysterious, and DI-2 remains dubious.

1.3. Acquiring a Belief by Finding One's Current Evidence Convincing on One's Current Evidential Standards

Let's try once more to specify the kind of belief-acquisition that's supposed to be necessarily non-intentional—the kind the involuntarist is labeling “belief-formation.” This time we might hold fixed not just the subject's body of evidence for the belief she is to form, but also the evidential standards she brings to bear on the evidence. If we alter DI to match that understanding of belief-formation, we'll be considering this proposal:

*DI-3: It is impossible to acquire a belief by finding your current evidence convincing on your current evidential standards, as an intentional action.*³

Finding evidence convincing, *without* changing your evidential standards: is that what belief-formation is? Is that what the involuntarists were getting at? I doubt it. Given certain (contested) assumptions about belief, DI-3 does seem to emerge as a metaphysical constraint on belief-acquisition. However, I argue, it turns out to be an uninteresting sort of constraint. If we like, we can certainly define a category of belief-formation with the characteristics posited by this version of DI, but doing so won't show that our doxastic lives are importantly non-intentional. The involuntarist who limits herself to endorsing DI-3, then, will be defending only a hollowed-out version of her position.

First, why would DI-3 be true? Well, suppose we allow, as a premise, that it's impossible both to believe a proposition *p*, and to believe that it's irrational for you to believe *p*. Suppose, further, that evidential standards are something like higher-order beliefs—beliefs about which beliefs your evidence supports. Given these two suppositions, your evidential standards are going to limit the beliefs it's possible for you to form. For instance, maybe your evidential standards require agnosticism about whether the number of stars in the Milky Way is even: that is, you have the higher-order belief that your evidence doesn't support a belief about that one way or the other. If you were able nonetheless to believe it's an even number, without getting any further

³ Special thanks to Sinan Dogramaci, here, for raising this possible interpretation of DI.

evidence, then you could have a belief that you yourself regard as unsupported by the evidence—that is, one you regard as irrational. But that’s what we’re supposing is impossible, and so we’ve reached a contradiction. If our assumptions are right, it seems you can’t *but* believe what your evidential standards prescribe.

At any rate, that’s an argument we could imagine someone making. I won’t stop to ask whether it’s a good one, since I’m not bothering about whether DI-3 is true or false. Instead I’m arguing that, whether true or false, the thesis constitutes only a trivial form of involuntarism. It fails to put a meaningful constraint on our powers of intentional action.

What do I mean by that? Imagine yourself in a situation like the one Susanna Rinard (2019: 1933–34) constructs: you and your mining team are trapped in a cave, and a rich eccentric—that indispensable figure—will have you rescued if and only if you believe the number of stars in the Milky Way is even.⁴ Imagine you’re talking to one of the other miners about the predicament. “It’d be great to acquire that belief, of course,” you say. “The problem is, I can’t acquire it while retaining my current evidential standards. So you see I’m really in a pickle.” Wouldn’t the miner look at you in angry consternation? Wouldn’t a reasonable response be: “Why on earth do you care whether those evidential standards stay the same? I can understand if you’re just *unable* to change them, psychologically. But if you *can*, then what are you waiting for? Don’t you realize lives are at stake, here?”

The moral I draw from that imagined scene is this: when we talk about whether we could form a belief intentionally, we’re interested in our ability to get the beliefs that have practical benefits for us. What we’re not particularly interested in is our ability to get beneficial beliefs without also altering our evidential standards or higher-order beliefs. Maybe we really do lack that latter ability. Maybe forming a belief does require holding an evidential standard that

⁴ Rinard puts the case to a different use than mine, looking at reasons for belief rather than at intentionality.

rationalizes it. But unless that claim is accompanied by a claim that your evidential standards can't be changed intentionally—a claim for which we haven't seen any support, so far, and for which section 1.2 has proposed counterexamples—it doesn't really bear on questions about our intentional control over belief.

It won't help her cause, then, if the involuntarist takes cues from DI–3 and just *defines* belief-formation as a process that keeps evidential standards fixed: so defined, belief-formation simply isn't an interesting classification. It would be hardly any worse to define “belief-formation” as “non-intentional belief-acquisition.” It'll be really easy to prove a version of DI governing *that*—but then, it won't be a version worth proving.

Part 2: Answering the Metaphysical Involuntarists

2.1. The Shape of the Problem

In Part 1, I was asking what it is, exactly, that DI claims can't be an intentional action: what is belief-formation? The answers I considered were unsatisfying. The situation seemed to be like this: the category of belief-formation could be demarcated *either* in such a way that it seemed meaningful in our cognitive lives (DI–1 and DI–2), *or* in such a way as to avoid strong counterexamples (DI–3)—but we couldn't have both. Apparently we had to choose: would our preferred form of involuntarism be false, or trivial?

I turn now to four arguments for metaphysical involuntarism, where we'll see the lessons of Part 1 repeated. Each of these arguments, I'll suggest, is plagued by a problem like the ones that afflicted our readings of DI. Three of them—Setiya's, Hieronymi's, and Benbaji's—address a cognitive process of real importance, but fail to establish their conclusions. The conclusion of Shah and Velleman's argument, on the other hand, is trivially weak in the manner of DI–3.

2.2. *Shah and Velleman's Trivially Narrow Deliberation*

Let's begin with that one. Shah and Velleman (2005) give an argument that may be sound, but it ends with surprisingly thin results. I'd formalize it like this:

General Principles:

- SV1. It is part of the concept of believing that its norm is truth.
- SV2. To deliberate about whether to φ is to engage in reasoning aimed at issuing or not issuing in φ -ing in accordance with the norms for φ -ing.

Derivation of Involuntarism:

- SV3. To deliberate about whether to believe that p is to engage in reasoning aimed at issuing in believing that p if and only if p is true (from SV1, SV2).
-

- SV4. It is impossible to deliberate about whether to believe that p , and to be moved by reasons you don't regard as bearing on the truth of p (from SV3).

Shah and Velleman begin from the idea that belief is partly constituted by having the standard or norm of truth, so that a belief is correct just in case it's true (SV1). This is supposed to be built into the very idea of belief. They further propose (SV2) a particular understanding of deliberation: they say it is "reasoning that is aimed at issuing in some result in accordance with norms for results of that kind" (2005: 502). These two moves allow them very efficiently to rule out the possibility of deliberately forming beliefs for non-truth-related reasons, as they proceed to do in SV3 and SV4. The nature of deliberation has neatly bracketed those reasons out.

Although Shah and Velleman don't explicitly connect their conclusion with the idea of intentional action, there's a way of understanding their argument in those terms. They're saying that practical reasons, because they're irrelevant to the norm of truth, can play no role in

deliberation about what to believe. But intentional action, whatever precisely it is, seems to require some sensitivity to practical reasons (see Bennett 1990: 90). If you can't respect these reasons while engaged in doxastic deliberation, then plausibly its results can't be determined intentionally. We can read Shah and Velleman, then, as giving a kind of argument for DI. They show that it's impossible to form a belief as an intentional action—if we understand “form a belief” to mean “acquire a belief as a result of doxastic deliberation.”

I'll dispute nothing in this argument. My complaint is that, like the argument for DI-3 above, it establishes only a trivial form of involuntarism. It's trivial because Shah and Velleman's “deliberation” is so narrowly drawn as to disappear as an important category, and a type of belief-formation defined as requiring such deliberation will inherit that insignificance.

Notice how their conclusion, like DI-3, seems strangely beside the point in the case of the trapped mining team. Imagine saying mournfully to the miner you're with: “I'd like to acquire the weird belief the eccentric billionaire wants. But when I deliberate about whether to believe it—in a certain special sense of the word ‘deliberate’—I can only be influenced by epistemic reasons. Ah, cruel world!” The miner won't be any more impressed than before. The response will be: “Look, if that kind of ‘deliberation’ won't let you take into account the fact that we're all in terrible danger, then maybe you ought to try just ‘determining’ or ‘considering’ what to believe, or doing whatever kind of thinking doesn't involve a hang-up about epistemic correctness.”

To put my criticism another way, suppose we accept that we can't deliberate about believing in the light of practical reasons. The question that remains will be: is there something that looks almost exactly *like* deliberating about what to believe, except that it doesn't involve a commitment to producing epistemically correct outputs? Shah and Velleman haven't ruled it out.

What they have established is, more or less, that it's impossible to intentionally "decide-for-exclusively-truth-related-reasons" whether to believe something, for non-truth-related reasons. I'm happy to agree, but that's not a serious limitation on the domain of intentional action. I thought doxastic involuntarism had more to it.

2.3. Setiya's *Atelic Progressives*

Setiya's argument for involuntarism (2008: 46–47) goes something like this:

General Principles:

- S1. If you are forming a belief as an intentional action, then you are intentionally forming a belief without regard for its truth.
- S2. If you are intentionally performing an action, then you believe you are performing it.
- S3. It is a conceptual truth that if you *are* forming a belief in some proposition, then you *have* become more confident in it.
- S4. It is a conceptual truth that it's irrational to become more confident in a proposition without regard for its truth.
- S5. It is impossible to have a certain level of confidence in a proposition, while also believing that this level of confidence is irrational.

Derivation of Involuntarism

- S6. If you form a belief as an intentional action, then you believe that you're forming a belief without regard for its truth (from S1, S2).
 - S7. If you form a belief as an intentional action, then—barring conceptual confusion—you believe that you *have* become more confident in some proposition without regard for its truth (from S3, S6).
 - S8. If you form a belief as an intentional action, then—barring conceptual confusion—you believe that your level of confidence in some proposition is irrational (from S4, S7).
-

- S9. Barring conceptual confusion, it is impossible to form a belief as an intentional action (from S5, S8).

S1 expresses Setiya's view of what intentional belief-formation would have to be (2008: 42). S2 is the application of an Anscombean principle of practical knowledge (Setiya 2008: 45; cf. Anscombe 1957/2000: 13–15), and S4 and S5 are unargued but (Setiya thinks) compelling theses about belief (2008: 42–43). I'll take all that for granted. S3 needs a little more explaining. First, the transition between “forming a belief” and “becoming more confident” is supposed to be painless: those are just the same thing here, we're told (2008: 45). As for the change of tense—from “you are becoming more confident” to “you have become more confident”—Setiya justifies it with linguistic principles. He claims that “become more confident that *p*” is, in linguistic terms, an “atelic” predicate, a verb phrase indicating an action with no natural endpoint or “target state” (Setiya 2008: 46; see Borik 2006: 23–55). The linguists have it that, where an atelic predicate is concerned, one can always infer from the present progressive form to the present perfect (Borik 2006: 26). Thus “I am walking” implies “I have walked,” and “she is singing” implies “she has sung.” So also, Setiya argues, “you are becoming more confident in *p*” implies “you have become more confident in *p*.” In contrast, “you are driving to Houston” doesn't imply “you have driven to Houston”; that's a “telic” progressive instead.

If all these premises were true, then we'd have a solid argument that intentional belief-formation is possible only for a confused agent. However, I reject S3. It's not guaranteed by linguistic rules, and it's just not a necessary truth, that whenever one *is* becoming more confident in something, one also *has* become more confident in it: Setiya has misapplied that “atelic progressive” label. Remember, atelic predicates are those that have no natural goal or

target—but “become more confident” *does* aim at a target state: namely, the state in which the agent is more confident than she was before.

Things are even clearer if we consider the phrase we have mainly been concerned with, which is not “becoming more confident” but “forming a belief.” That’s is a telic progressive if ever there was one. The natural endpoint of that activity is the existence of that belief, as the building of a sand castle aims at the existence of the sand castle. For these telic verb phrases, there is no entailment from the present progressive to the perfect. If a child is building a sand castle, it doesn’t mean she has already built a sand castle. And it could also be true that I’m forming a belief in *p*, without it being true that I have formed a belief in it.

Let’s return, though, to Setiya’s language of “becoming more confident.” We can table the question of telicity and simply ask: is it true that whenever anyone *is becoming* more confident, she must also *have become* more confident? Why should we accept this? It certainly isn’t true of “becoming” in general. If I tell you my friend *is becoming* a doctor, I haven’t implied that she *has become* a doctor; it could be years and years before that’s true.

What about when our verb phrase has a comparative adjective in its complement, as with “becoming more confident”? When we use phrases like that in the present progressive, admittedly, the present perfect tense is usually licensed as well. If I say that someone is becoming poorer, or taller, or more virtuous, you can generally assume she has advanced some of the way toward poverty, height, or virtue already—otherwise, what would have prompted my assertion, and what would justify it? But this isn’t a conceptual requirement, as it would be for atelic predicates. The connection between the present progressive and the present perfect can be broken when the prompt and justification of a claim are provided in some other way. For example, maybe you and I are watching a war hero receive a new medal. During the ceremony I

marvel aloud: “He’s already the most decorated American alive, and here he’s becoming even more decorated.” If the ceremony hasn’t yet reached the part where they actually award the medal to the war hero, then he hasn’t *already* become more decorated than he was before. Nevertheless, it’s fine to say he *is becoming* more decorated: the news we’ve heard about him, and other facts taken for granted in the context, have left us in no doubt that this process is going on. Can’t the same sort of thing be said for forming beliefs or becoming more confident? I’ve seen no reason to deny it.

The upshot is that I don’t believe there’s any such conceptual truth as the one S3 urges; and without that premise, Setiya’s argument collapses. It didn’t fail in ambition, though. If Setiya had shown that belief-formation (as he understands it) couldn’t be an intentional action, that really would have been a robustly involuntarist result. In his usage, forming a belief seems to involve no special, tacked-on requirements that acquiring a belief lacks. As with the arguments from Hieronymi and Benabji, below—and unlike with the one from Shah and Velleman—the problem here isn’t triviality, but unsoundness.

2.4. *Hieronymi’s Intentional Commitments*

I’d put Hieronymi’s (2009) argument in this form (after replacing some of her language with the vocabulary I’ve been using):

General Principles:

- H1. As soon as you intend to perform an action, you are committed to that action and all its intended parts.
- H2. To form a belief is to become committed to the truth of some proposition *p*.

H3. If you form a belief as an intentional action, the belief-formation is caused by the intention to perform it.

H4. Effects come after their causes, temporally.

Derivation of Involuntarism:

H5. As soon as you intend to form a belief that p , you are committed to the truth of p (from H1, H2).

H6. If you form a belief as an intentional action, then your commitment to the truth of p is caused by your intention to form the belief that p (from H2, H3).

H7. If you form a belief as an intentional action, then your commitment to the truth of p arises after your intention to form the belief (from H4, H6).

H8. If you form a belief as an intentional action, then your commitment to the truth of p both arises after your intention to form the belief, and exists as soon as the intention does (from H5, H7).

H9. Your commitment to the truth of p *cannot* both arise after your intention to form the belief, and exist as soon as the intention does.

H10. It is impossible to form a belief as an intentional action (from H8, H9).

H1 is meant to be a plausible principle of intentional action. As soon as you intend to make soup, for example, you are *committed* to making soup—or, committed to a positive answer to the question *whether to make soup*. When an intended action is non-basic, as soup-making is, you are similarly committed to each action involved in it: to dicing the onions, getting a pot from the pantry, and so on (2009: 159). This kind of commitment is required for an intention, so the intention can't exist without it even for a moment. Beliefs, meanwhile, come along with a different kind of commitment, the kind indicated by H2. When you believe some proposition p , you're committed to the truth of p , committed to a positive answer to the question *whether p* (2009: 160). Hieronymi also thinks an intentional belief-formation would have to be *caused* by an intention to believe (H3)—let's take that action-theoretic assumption on board, too. Finally,

because of H4, this means that an intentional belief-formation would have to begin *after* the intention it executes.

The argument Hieronymi makes from these premises takes some mulling over. Try thinking of it as a *reductio*. Posit a case where I intentionally form a belief. In such a case, my commitment to the truth of *p* (see H2) must result from my intention to form the belief (by H3); and so (by H4) my commitment to the truth of *p* comes *after* my intention to form the belief. But when I intend to perform an action, my commitment to the action and all its parts must arise *as soon as* my intention (by H1). Since my commitment to the truth of *p* arises *afterward*, that commitment must not really have been part of the action I intended. And since we could make the same argument for every putative doxastic intention, the belief-commitment can never be part of an intentional action at all.

That's my best interpretation. Now I turn to my refutation: at least as I have formulated it here, the argument is invalid. The mistake it contains is obscured by the multiple kinds of commitment under consideration (practical and theoretical), but if we look closely we'll detect it in the inference to H5 from H1 and H2: it does not follow from those premises that your belief, your commitment to *p*, must arise as soon as your intention to act. That would be so only if *p* itself—that is, some *proposition*—were an intended part of the action of forming a belief in it. In that case, because of H1, a commitment to *p* would indeed have to be present as soon as the intention was. But no one is arguing that the proposition itself could be an intentional action or a part of one. That would be a bad category error.

Here is how we should put things, instead, if we want to imagine intentional belief-formation on Hieronymi's terms. From the first moment I intend to form the belief, I must be committed to *making a commitment* to the truth of *p* (H1). That is plausible. In addition, the

making of that commitment—but not, notice, my *commitment* to making that commitment—must causally result from the intention to form the belief (H3); therefore the making of the commitment must occur after the intention to make it (H4). Also plausible. But these two plausible results do not contradict each other. The commitment to forming a belief in p must arise along with the intention to do so; and the commitment to p 's truth itself must arise afterward. No problem. Hieronymi's *reductio* leads to no *absurdum*.

An analogous case of intentional promising can illustrate this point. Like beliefs, promises carry a certain kind of commitment with them: normally, promising to φ entails becoming interpersonally committed to φ -ing. As soon as you intend to promise to φ , then, you are committed to making a commitment to φ . The promissory commitment itself, however—say, your commitment to help a friend move—needn't arise at the same moment as your intention to promise. Rather, it arises when you *execute* the intention, when you actually promise to help your friend. This delay between your commitment to making the promissory commitment and your actually making it in no way prevents you from being able to promise intentionally. Likewise, a delay between your intention and your belief doesn't show that belief-formation wasn't the intention's object.

2.5. Benbaji's "Because It Is So" Requirement

Finally, consider Benbaji's (2016) argument:

General Principles:

- B1. It is impossible to have a belief without regarding that belief as caused by what it represents.

- B2. If you are forming a belief as an intentional action, then you know that the belief is caused by your intention.
- B3. It is impossible *both* to know that a belief is caused by your intention, *and* to regard it as caused by what it represents.

Derivation of Involuntarism:

- B4. If you are forming a belief as an intentional action, then you *both* know that the belief is caused by your intention, *and* regard it as caused by what it represents (from B1, B2).

- B5. It is impossible to form a belief as an intentional action (from B3, B4).

The argument has a lot in common with the ones from Setiya and Hieronymi. Like Setiya, Benbaji features the non-observational knowledge of intentional action as a key premise; and like Hieronymi, she treats intentions as the causes of intentional actions. Those moves are represented together in B2: as Benbaji says, “if I could consciously believe at will . . . I would know that my forming the belief is brought about by my intention” (2016: 1954). As in the preceding arguments, these claims about intentional action are accompanied by a claim about belief, as expressed in B1: a belief must be conceived “as caused by what it represents, namely, the fact that *p*” (2016: 1954). Finally, here as before, all these characteristics of action and believing are deemed incompatible. The practical knowledge of action conflicts with the truth-responsive nature of believing, because “an attitude that is conceived to be caused by intention alone cannot also be conceived as caused by the fact it purports to represent” (2016: 1955). So says B3; and from there the impossibility of intentional believing comes along without too much prodding.

The biggest problem with Benbaji’s argument is yet another thing it shares in common with Setiya’s and Hieronymi’s arguments: it takes the feature thought to be a characteristic of

belief, and tries—unpersuasively—to project it back upon the action of belief-*formation*. Setiya tried to say that anyone becoming more confident in *p* must already have become more confident in it; Hieronymi tried to say that the commitment to *p*'s truth would have to exist as soon as the intention to form a belief in *p*. With Benbaji, the relevant feature of belief is the “because it is so” requirement, and the crucial move comes in the transition from B1 to B4. B1 asserts that, when we have come to believe something, we must regard our belief as caused by the fact it represents (2016: 1954). By the time we get to B4, though, this has become a requirement even on the *process* of belief-formation.

I don't mean to say this is a mere mistake in the argument's progression: projecting the “because it is so” requirement backwards onto the belief-formation is Benbaji's deliberate strategy. She asserts that the requirement doesn't simply govern our “retrospective knowledge about the origin of a belief,” but is instead “a condition on the very activity of *forming a belief*” (2016: 1958, original italics). But is there any reason to accept this maneuver? It can't be allowed as a general rule: there must be at least some necessary conditions on belief that won't be satisfied by belief-formation. (At minimum, *having the belief* is required for having the belief; but it can't be required for the process of forming of the belief, since once you actually have the belief the belief-formation will be over and done with.) So why should we think that this particular necessary condition for belief is also necessary for belief-formation?

Benbaji says little to answer that question. She does say, in a scoffing tone: “As if in forming a belief we can commit ourselves practically to the utility of the belief but not to its truth” (2016: 1958). Maybe there is, in fact, an ambiguity to be resolved here. It's true that belief-formation, considered as a whole and in a successful case, can't avoid committing one to the truth of a belief: without that, there hasn't really been a belief-formation at all. Forming a

belief is becoming committed. However, Benbaji's argument needs it to be the case that, at every moment *of* the intentional belief-formation, one is *already* committed to the belief's truth, or its causal origin in the fact that *p*—and that's what we haven't seen reason to accept. The building of a sand castle, considered as a whole and in a successful case, has to result in there being a sand castle. All the same, a child can be building a sand castle without there being one already. That's half the fun of it.

Conclusion

Time to tally up results. Among the four arguments for doxastic involuntarism in Part 2, I count one trivial conclusion and three unsupported ones. Even if we take for granted substantial claims about belief's nature and that of intentional action, we find the arguments for doxastic involuntarism vitiated in the same ways as its formulations in Part 1. Wherever we've looked, in fact, robustly involuntarist positions have seemed insupportable, while the positions secured by compelling arguments have seemed, themselves, more or less beside the point. Nowhere have we found a kind of belief-formation that is both important and impossible to do intentionally. Moreover, these problems weren't restricted, in their scope, to the forming of beliefs with any special content (see Peels 2015) or apparent evidential status (see Roeber forthcoming). It's true that intentional belief-formation isn't usually as easy as it is in Feldman's light switch case. But we've seen no principled reason that beliefs of *any* stripe would be *impossible* to form intentionally, as DI says they are.

I have to admit that there could be better readings of DI than the four I assessed, and there might be better ways of supporting them. I can only say, let's hear them! If my challenge only provokes involuntarists to answer it—to give an illuminating account of

belief-formation—I'll be satisfied. For now, I'll keep working on my doxastic passion projects. If I apply myself diligently, and if biotechnology advances at a brisk clip, maybe I really *can* form a belief that Mister Rogers likes me. Maybe I can do it intentionally.

References

- Alston, William (1988). "The Deontological Conception of Epistemic Justification," *Philosophical Perspectives* 2: 257–99.
- Anscombe, G. E. M. (1957/2000). *Intention*. Cambridge, MA: Harvard UP.
- Audi, Robert (2015). *Rational Belief: Structure, Grounds, and Intellectual Virtue*. Oxford: Oxford UP.
- Benbaji, Hagit (2016). "What Can We Not Do at Will and Why," *Philosophical Studies* 173: 1941–61.
- Bennett, Jonathan (1990). "Why Is Belief Involuntary?" *Analysis* 50/2: 87–107.
- Borik, Olga (2006). *Aspect and Reference Time*. Oxford: Oxford UP.
- Feldman, Richard (2000). "The Ethics of Belief," *Philosophy and Phenomenological Research* 60/3: 667–95.
- Hieronymi, Pamela (2009). "Believing at Will," *Canadian Journal of Philosophy*, suppl. vol. 35: 149–87.
- McHugh, Conor (2017). "Attitudinal Control," *Synthese* 194: 2745–62.
- Nottelman (2006). "The Analogy Argument for Doxastic Voluntarism," *Philosophical Studies* 131: 559–82.
- Pascal, Blaise (1670/2003). *Pensées*. Trans. W. F. Trotter. Mineola, NY: Dover.
- Peels, Rik (2015). "Believing at Will Is Possible," *Australasian Journal of Philosophy* 93: 524–41.
- Rinard, Susannah (2019). "Equal Treatment for Belief," *Philosophical Studies* 176/7: 1923–50.
- Roeber, Blake (forthcoming). "Evidence, Judgement, and Belief at Will," *Mind*.
- Setiya, Kieran (2008). "Believing at Will," *Midwest Studies in Philosophy* 32: 36–52.

Shah, Nishi, and J. David Velleman (2005). "Doxastic Deliberation," *The Philosophical Review* 114/4: 497–534.

Steup, Matthias (2012). "Belief Control and Intentionality," *Synthese* 188: 145–63.