

# UCLA

## UCLA Previously Published Works

**Title**

Responsibility for believing

**Permalink**

<https://escholarship.org/uc/item/9j39q7v0>

**Journal**

Synthese, 161(3)

**ISSN**

1111-4924

**Author**

Hieronymi, Pamela

**Publication Date**

2008-04-01

**DOI**

10.1007/s11229-006-9089-x

Peer reviewed

## Responsibility for Believing

Pamela Hieronymi  
hieronymi@ucla.edu  
May 3, 2006

To some it seems that rejecting doxastic voluntarism (rejecting the claim that we can believe at will) calls into question our responsibility for our beliefs. The underlying assumption seems to be that we are responsible only for what is voluntary. Thus, if we cannot believe voluntarily, we cannot be responsible for our beliefs. Yet it also seems clear that we are responsible for our beliefs: not only can we be charged with believing irrationally or unjustifiably, but we can also be, in our believing, respectful, generous, honorable, despicable, spiteful, or untoward. Moreover, we seem to be responsible for our beliefs in a rather direct way: we seem to be responsible for our beliefs, not merely as features of ourselves that we can manage or affect through our actions (not merely in the way we are responsible for, say, our health) but rather in something like the way we are responsible for our actions—despite the fact that believing is not voluntary.

Any tension here is only apparent. The underlying assumption, that we can be responsible only for what is voluntary, is false.<sup>1</sup> It holds in some central cases—and thus

---

<sup>1</sup> Robert Adams argues against this assumption, providing counter-examples, in “Involuntary Sins,” *Philosophical Review* 94 (1985): 3–31. For a more recent argument against the assumption, see Angela M. Smith, “Responsibility for Attitudes: Activity and Passivity in Mental Life,” *Ethics* 115 (January 2005): 236–71. The assumption seems to appear in the opening of Book III of the *Nicomachean Ethics*, where Aristotle says that feelings and actions “receive praise or blame if they are voluntary, but pardon, sometimes even pity, if they are involuntary.” Contemporary thinkers typically endorse this claim while remaining puzzled by Aristotle’s notion of voluntariness. I will try to articulate a more modern notion of voluntariness, arguing that this more modern notion is not required for responsibility.

It may be worth noting, at the start, that some thinkers would include, in the class of “the voluntary,” not only voluntary activities but also those things that are the (possible) result of such activity. I think it clear that believing can be the result of voluntary activity, and even, in the right circumstances, the product of voluntary activity aimed at bringing about that very belief. When it is, and *insofar as* it is, our responsibility for it should resemble our responsibility for any other state of affairs brought about by our voluntary activities, and so be unproblematic. Again, the special problem with responsibility for belief arises because it seems that we are somehow more directly responsible for our beliefs—responsible for them in a way beyond the way in which we are responsible for the management of our allergies or the state of our apartment—even though they are not themselves voluntary activities (not, as it is sometimes put,

acquires its plausibility—but it does not hold generally. In fact, on at least one plausible account of what it is for a thing to be voluntary and what it is to be responsible for something, beliefs are not voluntary and yet, for failing to be voluntary, they are a central example of the sort of thing for which we are most fundamentally responsible. This paper sketches such an account.

The first section examines the notion of responsibility. It starts with a very minimal notion, which I call *answerability*. It seems plausible that we are “responsible” for our beliefs in at least the minimal sense that we can rightly be asked for our reasons for believing (which is not to say that we must have reasons for each of our beliefs). This minimal notion is then shown to be continuous with a more robust notion of responsibility.<sup>2</sup>

The second section turns to voluntariness. It proposes an account of voluntariness on which ordinary actions are voluntary but believing could not be. Indeed, on the proposed account of voluntariness a wide range of important attitudes (“commitment-constituted attitudes”) could not be voluntary.

The final section brings together the account of responsibility and the account of voluntariness proposed, pointing out that the same features which make an attitude essentially non-voluntary, on the proposed account of voluntariness, also make it the sort of thing for which one is responsible, on the proposed account of responsibility.

---

“directly” voluntary). I will be concerned with this special problem. (I hope these matters will become clearer below.)

<sup>2</sup> Throughout, I take the question at hand—whether or why we are *responsible* for our beliefs—to be asking whether or why we are *open to* assessment or appraisal or praise or blame or credit or discredit on account of our beliefs. I do not take up the distinguishable question of what it is to believe “responsibly,” that is, what must be true if the assessment of one’s beliefs is to be positive.

Believing, then, is a prime example of the sort of thing for which one is responsible, due to features that also guarantee that it could not be voluntary. The remainder of the paper suggests how the original assumption, that one can be responsible only for the voluntary, gained its plausibility.

#### RESPONSIBILITY: AN ACCOUNT

‘Responsibility’ means different things to different people. I suggest starting with a very minimal notion, which I will call *answerability*. One is *answerable* for an activity or a state of affairs just in case one can rightly be asked for the reasons, if any, for which one engaged (or engages) in the activity or brought about the state of affairs.

I draw here from G. E. M. Anscombe, who noted that whenever one intentionally  $\phi$ 's (where  $\phi$ -ing stands for some ordinary action, such as doing the dishes or pumping water to the house), one can rightly be asked, “Why did you  $\phi$ ?” (or “Why are you  $\phi$ -ing?”) where this question looks, not for an explanation of how it came about that one  $\phi$ -ed, but rather for a the agent's reasons for  $\phi$ -ing.<sup>3</sup> This why-question is, in Anscombe's terms, “given application” whenever one acts intentionally.

---

<sup>3</sup> See G. E. M. Anscombe, *Intention* (Oxford: Basil Blackwell, 1957): 9. (The account of answerability I will present draws on, but is not wholly faithful to, Anscombe's.) Anscombe describes the sense of the question ‘Why?’ as “that in which the answer, if positive, gives a reason for acting” (9). A reason for acting, in Anscombe's usage and mine, is not (necessarily) a psychological state; it might be a non-psychological fact or consideration that the agent takes to count in favor of action, or to show something desirable. These non-psychological considerations will have some role in an explanation of the agent's action. As Donald Davidson puts the point (he and Anscombe here stand in agreement), “We cannot explain why someone did what he did simply by saying the particular action appealed to him; we must indicate what it was about the action that appealed,” that is, we must indicate “some feature, consequence, or aspect of the action that the agent wanted, desired, prized, held dear, thought dutiful, beneficial, obligatory, or agreeable.” This feature, consequence, etc., Davidson calls “the agent's reason for doing what he did” (as opposed to what Davidson will later call the “primary reason” for the action). (See “Actions, Reasons, Causes” in *Essays on Actions and Events* (Oxford: Oxford University Press, 1980): 3.) Neither Davidson nor Anscombe think the agent's reason—the feature, consequence, etc.—needs itself to be a psychological state. It may be, e.g, the illumination of the room. Davidson thought that the feature, etc., should appear in the content of psychological states that explain the action (e.g., a desire to illuminate

Importantly, such questions are given application by more than intentional action. A similar question is given application by belief: whenever one believes that  $p$  (where  $p$  stands for a proposition, such as “The butler did it” or “It is going to rain”), one can rightly be asked, “Why do you believe  $p$ ?” where that question looks, not for an explanation of how it came about that one believes, but rather for considerations that one takes to bear positively on whether  $p$  (that is, roughly, one’s reasons for believing). One’s belief will be justified or not (on one understanding of justification), or reasonable or not, depending on the quality of the answer one can give to this question. Further, if one can be brought to see the inadequacy of one’s answer, one will then (insofar as one is rational) abandon the belief.<sup>4</sup> Given these facts about justification, reasonableness, and belief revision, it seems natural to think that being answerable for our beliefs is one way—a minimal way—in which we are responsible for them. Even this minimal sense of responsibility, *responsibility as answerability*, allows one to be open to assessment as justified or unjustified, reasonable or unreasonable, in one’s believing.

Consider next the continuity between responsibility as answerability and more familiar features of “moral” responsibility.

To this end, reconsider Anscombe’s claim that a certain why-question is *given application* by intentional action. It is important to understand what she means by “given

---

the room). This last claim seems broadly correct to me (I consider it in “Reasons for Action” (ms.)), but nothing in this paper hinges upon it. For present purposes it is only important to understand what would satisfy the relevant why-question. It will be satisfied by the agent’s own reasons for  $\phi$ -ing (if any), that is, those considerations, if any, that the agent, in some sense, took to count in favor of  $\phi$ -ing. The relation of these reasons to the explanation of the agent’s  $\phi$ -ing is not presently at issue. I am grateful to Nikolaj Nottelman for pressing me to clarify these issues for the reader.

<sup>4</sup> “Adequacy” here is broadly understood. The answer, “There is a reason, but I can’t remember it,” or “there must be a reason,” or “I just think it is true,” or “it just seems that way,” might be adequate, in certain contexts. Further, even when such an answer is inadequate, whether one abandons the belief will turn on one’s own assessment of the adequacy of that answer (and one’s rationality).

application.” She says, “the question [‘Why did you  $\phi$ ?’] is not refused application because the answer to it says that there is *no* reason, any more than the question how much money I have in my pocket is refused application by the answer ‘None’.”<sup>5</sup> Rather, the question is refused application (according to Anscombe) by the answer, “I didn’t know I was  $\phi$ -ing,” just as (presumably) the question “How much money do you have in your pocket?” would be refused by the answer “I have no pockets.” The latter question is refused because an assumption made in asking the question—that you have a pocket—is shown false. It seems, then, that a question is given application just in case the assumptions made in asking it are met.

What, then, is the assumption that gives Anscombe’s question application? It cannot be the assumption that the person *had a reason* for  $\phi$ -ing: that assumption would be false, and so the question refused, if one  $\phi$ -ed for no reason. I suggest that it is rather the assumption that the person has, in some sense, settled (positively) the question of whether to  $\phi$ . To  $\phi$  for certain reasons is to have settled the question of whether to  $\phi$ , for those reasons.<sup>6</sup> Of course, if you have settled a question, then it makes sense to ask you for the reasons—if any—for which you have done so. Anscombe’s why-question simply inquires after those reasons.<sup>7</sup> The question “Why did you  $\phi$ ?” is thus given application

---

<sup>5</sup> Ibid, 25.

<sup>6</sup> As should become clear, the appeal to “settling a question” is not meant to introduce an additional psychological state or event. Rather, the claim “to intend to  $\phi$  is to settle positively the question of whether to  $\phi$ ” simply notes the uncontroversial conceptual connection between an intention and a positive answer to a certain question (whether to  $\phi$ ).

<sup>7</sup> Anscombe herself allows a slightly wider class. I would have to treat the exceptions piecemeal, hoping to show that they prove the rule.

by the truth of the assumption that the person  $\phi$ -ed (or is  $\phi$ -ing) because he or she has settled the question of whether to  $\phi$ .<sup>8</sup>

The insight can be extended: any activity or attitude for which one is answerable—for which one can be asked one's reasons—will be reasonably understood as (or as the result of) the settling of some question (or set of questions) on which such reasons bear. For example, one can be asked for one's reasons for believing  $p$ , and believing  $p$  can be understood as settling (or having settled) for oneself (positively) the question of whether  $p$ <sup>9</sup>—the question on which one's reasons for believing  $p$  will bear. We have already seen that one can be asked for one's reasons for intending to  $\phi$ , and intending to  $\phi$  can be understood as settling (or having settled) for oneself (positively) the question of whether to  $\phi$ . One can be asked for one's reasons for  $\phi$ -ing just in case one's  $\phi$ -ing was intentional, that is, just in case it is the result of one's having settled the question (however implicitly) to  $\phi$ . Likewise, the states of affairs for which one is answerable can be understood as those that one intended to bring about. I am answerable for my reasons for making such a mess only if the mess can be understood as something that I brought

---

<sup>8</sup> The question is likewise readily refused application (as Anscombe claims it should be) by the claim that one did not know that one was  $\phi$ -ing: absent the possibility of unconscious decisions, the claim that one did not know one was  $\phi$ -ing undermines the assumption that one is  $\phi$ -ing because one settled the question of whether to  $\phi$ . See *Intention*, 25ff. (The ideas of this paragraph also appear in my “The Will as Reason” (ms.) )

<sup>9</sup> Again, the claim that “to believe  $p$  is to have settled the question of whether  $p$ ” should be uncontroversial. The appeal to “settling the question” is not meant to introduce a new mental state or event. By “settling the question” I do not mean, e.g., consciously entertaining the question. The mental state or event that corresponds to “settling the question whether  $p$ ” is nothing other than believing  $p$ . The claim simply makes explicit the uncontroversial conceptual connection between believing  $p$  and a positive answer to a certain question—whether  $p$ . Insofar as one agrees that one's beliefs represent the world as one takes it to be, and that the question whether  $p$  asks whether the world is as  $p$  would have it, then one should have no trouble with the claim that to believe  $p$  is to have settled for oneself the question of whether  $p$ .

about either by intending to make it or by intending to do something else, of which the mess was a foreseen consequence.<sup>10</sup>

Notice, though, that if this is right—if the attitudes for which one is rightly answerable for one’s reasons are those that can be understood as one’s settling the question(s) on which those reasons bear, and if the actions or states of affairs for which one is rightly answerable are those that are either the execution of or the intended or foreseen consequence of one’s intentions—then we can start to see a continuity between the minimal sort of responsibility we have thus far considered and more familiar features of “moral” responsibility.

To be *morally responsible* for a thing (whether an attitude, an action, or a state of affairs) is to be open to moral assessment or judgment on account of that thing, to be open to moral praise or blame for it, to be open to what have come to be called the “reactive attitudes” of resentment and indignation on account of it, and perhaps to be open to moral sanctions or punishments for it. Setting aside sanctions and punishments, consider the relation between, on the one hand, being *open to moral assessment, moral judgment, or the reactive attitudes* on account of an activity, attitude, or state of affairs and, on the other, being *answerable* for that activity, attitude, or state of affairs:

We have said that the activities, attitudes, and states of affairs for which one is rightly answerable reveal one’s answer to some question. But, in so doing, these activities,

---

<sup>10</sup> In claiming that one is answerable for the foreseen consequences of one’s intentional actions, I part company with Anscombe. (See *Intention*, §25.) It is also worth noting that here we see some distance between answerability and a more full-blooded account of moral responsibility. Although one is *answerable* for a state of affairs only if it was the intended or foreseen outcome of some decision, one can be *responsible* for a state of affairs when it is merely a reasonably *foreseeable* consequence of one’s actions—even if one did not in fact foresee it. (This is because one can be responsible for a state of affairs even when it is not the result of one’s actions or decisions at all, but rather simply something about which one has obligations, which one might meet, neglect, or violate.)



attitudes, and states of affairs will also reveal something of one's mind, of one's take on the world and what is important or worthwhile or valuable in it. By settling certain questions for oneself, by having a take on what is true, what is important, and what is to be done, one thereby constitutes those bits of one's mind relevant to the quality of one's relations with others—and so establishes what we might call one's *moral personality*, or, in an older but apt phrase, the *quality of one's will*.<sup>11</sup> But that bit of one's mind—one's moral personality or one's will (broadly construed)—just is the object of moral assessment and reaction. It is that which we assess, when we assess whether someone is morally praise- or blameworthy; it is that to which the reactive attitudes react.<sup>12</sup>

---

<sup>11</sup> The evocation of “the will” might mislead. Adams discusses the difference between an older and a more modern notion of the will. (See “Involuntary Sins,” 6–9.) On the older notion, one's will is one's “rational appetitive faculty,” and so encompasses (some of) one's desires or appetites. In a similar vein, Harry Frankfurt notes that one's decisions, though in some sense a product of one's will, may not wholly reflect one's will. He points out the “exaggerated significance [that] is sometimes ascribed to decisions, as well as to choices and similar ‘acts of will.’” If we consider that a person's will is that by which he moves himself, then what he cares about is far more germane to the character of his will than the decisions or choices he makes.” See “The Importance of What We Care about,” *The Importance of What We Care about* (Cambridge: Cambridge University Press, 1988): 84. Though I will be articulating and defending a more modern notion of “voluntary,” when I talk of “the quality of one's will,” I mean to appeal to something closer to this broader, older, notion of the will.

Further, by appealing to “quality of will” or “moral personality,” I am quite self-consciously avoiding reference to “character” or “virtue.” The latter terms tempt one to think in terms of dispositions, but thinking in terms of dispositions obscures what is central to the present discussion: the role of agency. So by “moral personality” or “will” I mean to refer to a collection of attitudes (those T. M. Scanlon calls “judgment-sensitive attitudes,” which I call “commitment-constituted attitudes”) each of which represents the subject's take on some object, the subject's answer to some question(s). These attitudes will be understood, in a rational creature, to interact in sensible and predictable ways, mutually adjusting, responding to the environment, explaining both each another and the subject's overt actions. That interaction, mutual adjustment, and response to the environment just is the subject's activity; it is her forming and revising her mind.

<sup>12</sup> One might insist that not *all* the attitudes for which one is answerable should be considered part of one's moral personality or thought to reveal the quality of one's will, pointing out, e.g., beliefs or intentions that are the result of phobias or hypnosis. While this is difficult terrain, for the purposes of this paper it is enough that those attitudes that do uncontroversially comprise one's moral personality are things for which one is answerable. This claim establishes continuity between answerability and moral responsibility. In the next section I will argue that attitudes for which one is answerable—those that represent one's own take on what is true, good, or important—could not possibly be voluntary, precisely because they represent one's own take on the world. Thus, those attitudes for which one is uncontroversially responsible could not be voluntary, and so voluntariness could not be required for responsibility. (If one wants to distinguish

To morally assess someone on account of some *action or attitude*—to say that she was, in that action or attitude, respectful or disrespectful of others, or to say, in light of that action or attitude, that she was being kind, generous, magnanimous, spiteful, petty, or cruel—is not just to make a statement about the action or attitude; it is to make a statement about the quality of the will or mind or moral personality of the which action or attitude is a product or part. It is to make assumptions about the mind from which the action emanated, or the mind into which the attitude fits, and to react to it—to that mind or moral personality.<sup>13</sup>

To morally assess someone on account of some *state of affairs*—to take the state of affairs as one’s ground for resentment, or blame, or gratitude, or for thinking that someone is petty or conscientious or brave—one needs to think that the state of affairs in some way reveals something of the person, of the quality of her will. If a state of affairs is one for which a person is answerable—if it was the intended or foreseen outcome of some her executing some intention—then it can reveal something of the quality of her will.

So the minimal way in which it seems most plausible to think that one is responsible for one’s beliefs—namely, one is answerable for an account of the reasons for which one believes—is thus continuous with more ordinary features of moral responsibility. Again, one is answerable for an account of one’s reasons for those things that can be understood

---

between those attitudes that are and those that are not properly considered part of one’s moral personality, one thus cannot do so by appeal to whether the attitude is voluntary.)

<sup>13</sup> This sort of assessment differs in important ways from the determination of whether a given action was, in the circumstances, morally permissible. Determinations of permissibility may part company from the particular moral personality of the agent, the quality of her particular will—it is possible to innocently do what is impermissible, or to do what is permissible with blameworthy motives. (There may be a parallel here with the fact that determinations of epistemic warrant sometimes part company with the reasons of the believer.)

as one's settling for oneself some question(s) or as the intended or foreseen consequence of executing some intention. But anything that can be so understood will reveal one's take on what is true or important or worthwhile, and so can help to determine the quality of one's will. But the quality of one's will just is the object of moral assessment and the reactive attitudes. The quality of one's will just is what is assessed, judged, or responded to, when one is morally responsible. So those things for which one is answerable will be things for which one can also be morally responsible—for which one can be open to moral assessment and reaction.<sup>14</sup>

Once we see how responsibility as answerability is continuous with the more familiar features of moral responsibility, we might want to construct a parallel sense of epistemic responsibility, one which is a bit more robust than the minimal notion of responsibility as answerability. On such an account, to be *epistemically responsible* for a thing would be to be open to epistemic assessment or judgment on account of that thing, to be epistemically praise- or blameworthy for it, to be open to certain reactions from others on account of it, and perhaps to be open to certain sanctions from one's epistemic community. The difference between "moral" and "epistemic" responsibility, on such an account, would simply be which standards of assessment are applied, and, correspondingly, which reactions are warranted.

---

<sup>14</sup> On the current understanding of responsibility, to be responsible for something is to be, on account of that thing, open to assessment and, depending on the outcome of that assessment, the appropriate target of certain responses. On such an understanding, one is responsible for things that are morally innocuous (intentionally dropping one's keys on the desk, choosing vanilla rather than chocolate ice-cream). One is *open* to assessment on account of them; the outcome of the assessment would be neutral, and no particular reactions would be warranted.

Thus we see that the minimal notion of responsibility as answerability is continuous with a more robust sense of responsibility, where to be (more robustly) *responsible* for something is to be open to (moral or epistemic) assessment or judgment on account of that thing and, depending on the outcome of that assessment (whether it is positive, negative, or neutral), the appropriate target of certain reactions on account of it.<sup>15</sup> The things for which one is answerable are things for which one can also be responsible in this more robust sense, insofar as the things for which one is answerable reveal one's answers to certain questions, and so can reveal one's mind, one's (moral or epistemic or rational) self, which is the object of both assessment and reaction.

#### VOLUNTARINESS: AN ACCOUNT

What, then, of voluntariness? Initially it seemed that one can be responsible for something only if that thing is voluntary, and yet it also seemed that we are responsible for our beliefs despite the fact that our beliefs are not voluntary. I have suggested that one can be responsible for a thing if one is answerable for the reasons (if any) for which does, did, or brought about that thing. If the voluntary were identified with the class of things that one can do or bring about for reasons, believing would be unproblematically voluntary. Since it seems implausible to claim that believing is voluntary, we must have something else in mind. Specifying exactly what this is proves somewhat difficult.<sup>16</sup>

---

<sup>15</sup> Cf. T. M. Scanlon's notion of "responsibility as attributability." See "Responsibility," *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998): 248–94.

<sup>16</sup> Adams attempted it in "Involuntary Sins," pp. 7–9. Before embarking on several pages of desiderata, he notes, "It is not a simple matter to explain . . . the decisive criterion of the voluntary for most modern thought." Jonathan Bennett seems to give a simpler characterization: "I understand voluntariness as responsiveness to practical reasons. I take  $\phi$ ing to be voluntary if . . . one can  $\phi$  or not depending on whether one thinks one has practical reasons to  $\phi$ ." (See his "Why is Belief Involuntary," *Analysis* 50 (1990): 90.) In fact, Bennett's understanding of practical reasons requires him to complicate this account considerably: if believing is voluntary it must be possible that Bennett could "without giving you evidence

When considering whether believing is voluntary (or why it is not), believing is often contrasted with things that can be done “just like that,” “by deciding to.”<sup>17</sup> Raising one’s right hand is often proposed as a contrast. But believing contrasts just as well with more complex actions, actions that cannot be done “just like that,” such as doing the dishes, rearranging the furniture, or driving across the country. Each of these are voluntary in a way that believing seems not to be. So we need to find a notion of voluntary that will carve between believing, on the one hand, and, on the other, both raising one’s right hand and rearranging one’s furniture.

Perhaps, then, we can understand voluntary activities as those that we can do by deciding or choosing to do them. Assuming we are able-bodied, we can raise our right hand by deciding to do so, and, again assuming that we are able-bodied, we can rearrange the furniture by deciding to do so. Of course, we cannot rearrange the furniture *simply* by deciding to. To get it rearranged, we will have to do a number of other things: lifting and pushing and pulling until we have accomplished our end—the new arrangement. And, of course, features of the world beyond our own body will (as it is sometimes put) need to cooperate. Still, there is something to the thought that we can rearrange the furniture by deciding to do it and so doing it (providing all goes well), just as we can raise our right hand by deciding to do it and so doing it (providing all goes well). Perhaps, then, we

---

for the belief and just be offering convincing bribes and threats, get you to acquire the belief ‘just like that,’ immediately, without setting your thought on producing some intervening event that will lead you to getting the belief.” (See pp. 88–89.)

<sup>17</sup> In a modern *locus classicus*, Bernard Williams claims, “it is not a contingent fact that I cannot bring it about, just like that, that I believe something, as it is a contingent fact that I cannot bring it about, just like that, that I am blushing.” In the context of Williams’s discussion, to “bring it about, just like that” that one does something is to do that thing at will. See “Deciding to Believe” *Problems of the Self* (Cambridge: Cambridge University Press, 1973): p. 148 (originally published in H. Kiefer and M. Munitz, eds. *Language, Belief, and Metaphysics* (Albany: State University of New York Press, 1970)).

cannot believe voluntarily because we cannot, even if all goes well, believe by deciding to believe.

There are (at least) two troubles with this thought. First, there is a ready sense in which we can believe by making a decision. In many cases, we can consider the evidence for or against  $p$ , come to a decision about whether  $p$ , and so believe  $p$ . We often decide whether we can make the movie on time, whether the paper deserves an “A,” or whether a given project is likely to succeed. To attempt the needed distinction, then, we must rule this out. We might add a restriction: perhaps an activity is voluntary if (providing all goes well) one can do it by making a decision *for practical reasons*. That is to say, perhaps an activity is voluntary if you can do it by making a decision, where you can decide to do it for any reason you take to count sufficiently in favor of doing it. You can decide to raise your right hand or rearrange the furniture for any reason you take to show it sufficiently worthwhile: you could rearrange the furniture because doing so would optimize your space, please (or to spite) your spouse, relieve (or to express) your boredom, make a good joke. In contrast, although believing  $p$  might please your spouse or relieve your boredom, and though you might take this to show believing  $p$  sufficiently worthwhile, it seems plausible to claim that you cannot, for these reasons, decide to believe and thus believe (even if all goes well).

Even thus restricted, the analysis runs into a second problem: once we allow that the intermediate activity—the lifting and pushing and pulling—does not disqualify rearranging the furniture as something we can do by deciding to it, one might wonder why believing is disqualified. Having decided to create a new arrangement of the furniture, we then do a number of different things (lifting and pushing and pulling) until

(providing all goes well) we have achieved our end. But it seems that something similar can be said about believing: in suitable circumstances, one can, for practical reasons, decide to create in oneself a belief that  $p$  and then undertake some process—making a phone call, conducting an investigation, taking a pill—that will (providing all goes well) have as its result that one believes  $p$ .<sup>18</sup> Once we notice that we can, providing all goes well, produce a belief in ourselves, why would we deny that we can believe  $p$  by deciding to, for practical reasons, in the same way that we can rearrange the furniture by deciding to, for practical reasons?

One might insist that we can rearrange the furniture by deciding to rearrange it, because *rearranging* is an action, while we cannot believe by deciding to believe, because *believing* is not an action. Rather, believing is, in certain circumstances, the possible *product* of actions aimed at bringing it about, just as a particular arrangement of the room is the possible product of actions aimed at bringing it about. The fact that we can bring about a belief, one might say, no more makes believing an action than the fact that we can rearrange the furniture makes the particular arrangement of the room an action. The actions at hand are rather rearranging the room (a complex action comprised of lifting, pushing, pulling, with the aim of producing the arrangement) and producing the belief (a complex action comprised of conducting the investigation, or making the phone

---

<sup>18</sup> The claim that one can, in suitable circumstances, create in oneself a belief that  $p$  might need some defense. Consider some mundane examples. If you want to produce in yourself a belief that the lights are off, the required process is not difficult: you can get up and throw the switch. (The example belongs to Richard Feldman, “The Ethics of Belief,” *Philosophy and Phenomenological Research* 60.3 (2000): 671–2.) If, unable to sleep, you want to believe that your children have arrived home, you can place a call. (The example belongs to Thomas P. Kelly, “The Rationality of Belief and Some Other Propositional Attitudes,” *Philosophical Studies* 110 (2002): 169–70.) If you want to believe that your friend has never betrayed you, you might take a pill that induces amnesia about the relevant stretch of shared history. Like furniture, a belief that  $p$  is the sort of thing that interacts in more-or-less predictable ways with its environment. We regularly make use of this fact in our (often successful) attempts to produce beliefs in others. It should not be surprising that, given suitable circumstances, we can devise ways to produce beliefs in ourselves.

call, taking the pill, with the aim of producing the belief). These actions can be done “by deciding to do them, for practical reasons.” Believing cannot.

This reply is correct, but not yet helpful. If we were to use this analysis to draw the desired distinction, we would end up saying that believing is not voluntary because believing is not an action. For this to be helpful, we need to understand *why* believing is not an action—we need to understand why activities which bring about a new arrangement count as the action of *rearranging the room* while activities which bring it about that one believes *p* count, not as believing, but only as the action of *bringing it about* that one believes.

Here is a suggestion: believing could not be understood as an action because believing brings with it its own distinctive form of answerability.<sup>19</sup> In believing, you are answerable for reasons that you take to show the belief true. In contrast, in either intending to act or acting intentionally (where the action may be as complex as you like), you are answerable for reasons that you take to show something good about so acting. If we were to try to make belief into an action, one would have to be, in believing, answerable for reasons that one takes to show something good about believing. But these are a different class of reasons than those one takes to show that *p*.<sup>20</sup> Because believing

---

<sup>19</sup> Someone might have tried to say that rearranging is an action while believing is not because the activities involved in rearranging all have a common end, the end of achieving the rearrangement, and this allows them to be grouped into the action of “rearranging.” Believing, though, is not end-directed, so the relevant action could not be “believing;” it must rather be “producing the belief”—producing the belief is the common end that allows the activities to be grouped into an action. This approach has much in common with the approach I prefer: the common end ensures that the activities share a single sort of answerability. My preferred approach, to appeal directly to answerability, has the advantages of both being able to be extended beyond the case of action (as will be seen below) and not requiring that all voluntary action be end-directed.

<sup>20</sup> I have much to say about the difference between these reasons (as well as the overlap between them) in “Controlling Attitudes” *Pacific Philosophical Quarterly* 87, no. 1 (March 2006): 45–74 and “The Wrong Kind of Reason,” *Journal of Philosophy* 102, no. 9 (September 2005): 437–57. A more detailed, somewhat different version of the argument of this section in appears in “Controlling Attitudes.”



thus entails its own distinctive form of answerability, believing  $p$  cannot be understood as an action in its own right. To put the point somewhat differently, because believing brings with it its own distinctive form of answerability, believing is not the proper object of an intention.

Consider a case. Suppose you find yourself with reasons that you take to show it good to believe  $p$ , which you do not also take to show that  $p$ . Since you do not take these reasons to show that  $p$ , you will not, straightaway, believe. However, you might, for these reasons, form an intention with respect to this belief—the supposed “intention to believe  $p$ .” Of course, to successfully execute this intention, you will have to bring it about that you have settled the question of whether  $p$ ; only so will you incur the appropriate form of answerability. But, of course, that is just to say, to successfully execute this intention, you will have to bring it about that you believe  $p$ . The intention, then, is best understood, not as an “intention to believe  $p$ ,” but rather as an intention *to bring it about* that you believe  $p$ , an intention to make it the case that you settle the question of whether  $p$  positively, and so believe. Bringing something about is an ordinary, voluntary action; believing cannot be.

Believing might be usefully compared with “activities” that cannot be done for reasons at all, or activities that are essentially unintentional, such as digesting your food, sneezing, or tripping unintentionally. We might talk about “intending to digest your food more slowly,” “intending to sneeze,” or “intending to trip unintentionally.” But in each case the intention, properly speaking, is an intention to bring about something that is not itself an intentional action—an intention *to bring it about* that you digest your food more slowly, sneeze, or trip unintentionally. I am suggesting that an “intention to believe”

must be understood similarly, as, properly speaking, an intention *to bring it about* that you believe—not because believing is not done for reasons (or is unintentional), but rather because it is done for a different kind of reason than those for which you might intend or act.

With this much understanding, I propose the following account of voluntariness: an activity is *voluntary* just in case you decide to do it for reasons you take to settle the question of whether to do it, therein intend to do it, and, providing all goes well, do it by executing that intention. That is, an activity is voluntary just in case one engages in the activity by forming and executing an intention to do so, where one forms an intention to engage in the activity by settling the question of whether to engage in it, where that question could be settled by any set of considerations that one takes both to count in favor of the activity and to be sufficiently strong (or, indeed, for no particular reason).

This account ignores many of the central uses of ‘voluntary,’ where saying one did something voluntarily means something like one did it without coercion or undue pressure or obligation. In fact, this account collapses the voluntary activities and intentional activities. While there is very good reason to resist this collapse, and to preserve the word ‘voluntary’ for other uses, it seems to me that this account captures what both the discussion of “doxastic voluntarism” and discussions of responsibility often mean by ‘voluntary:’ an activity is voluntary just in case (given the right background conditions, and providing all goes well) you do it by deciding to do it, where you can decide to do it for any reason you take to show it worth doing.

This account sorts the cases in the required way. Raising one’s right hand is voluntary, on this account: one raises one’s right hand by forming and executing an

intention to do so. One intends to raise one's right hand by settling the question of whether to raise it, where that question could be settled by any (set of) consideration(s) that one takes to count sufficiently in favor of raising one's right hand. Rearranging the furniture is also voluntary, on the proposed account: One rearranges the furniture by forming and executing an intention to do so. One forms an intention to rearrange the furniture by settling the question of whether to do so, where that question could be settled by any (set of) consideration(s) that one takes to count sufficiently in favor of rearranging.

In contrast, believing is not voluntary, on this account. One cannot believe  $p$  by settling the question of whether to believe  $p$ , where that question could be settled by any set of considerations that one takes to count sufficiently in favor of believing  $p$ . As we have seen, one might well take the fact that believing  $p$  would make a good joke, or relieve one's boredom, or spare one anguish, to count sufficiently in favor of believing  $p$ . But one cannot, for those reasons, settle the question of whether to believe  $p$ , therein form an intention to believe  $p$ , and execute that intention by believing  $p$ . This is because, when one believes  $p$  one is understood to have settled the question whether  $p$ , but one cannot be understood to have settled *that* question for reasons which one only takes to show believing good to do. And so one could not, for such reasons, become answerable in the way one must, if one is to believe. One will rather, for reasons that one takes to show believing good to do, settle the question on which one takes those reasons to bear—the distinct question of whether believing is good to do—and so form a second-order belief about the belief that  $p$ . In the right circumstances, one might then decide to bring it about

that one believes  $p$ . Bringing it about that one believes  $p$  could be an ordinary voluntary action. Believing could not be.

Importantly, believing is not the only attitude that is in this way not voluntary.

Resenting, caring, fearing, being grateful, being angry, are none of them voluntary, in this sense—and all for the same reason. Each of these attitudes, like belief, expresses its subject's take on, opinion about, or orientation to, some object. When one resents, or is grateful, or cares, or is angry, one is understood to have a certain take on the object of one's resentment, gratitude, care, or anger, a take for which one is answerable. One is understood, we might say, to have settled for oneself some range of questions about that object. But, just as we could not understand you as answering for yourself a question about whether  $p$  for reasons that you take only to show believing good to do, so we cannot understand you as answering the range of questions involved in resenting or being grateful or caring or being jealous or angry, for reasons that you take to bear only on whether the attitude is good to have. We could rather, in the right circumstances, understand you as *making yourself* angry, or resentful, or jealous, or bringing yourself to care, for these reasons (or prompting anger, etc., in yourself, for those reasons).<sup>21</sup>

So, a large and important range of attitudes are not voluntary. Elsewhere I have called these “commitment-constituted attitudes.”<sup>22</sup> Note, though, that while these attitudes are

---

<sup>21</sup> Some states of mind one can voluntarily adopt. One can, e.g., visualize world peace, suppose  $p$  for the sake of argument, or bisect a triangle in one's imagination, by deciding to do so. One can decide to do so for any reason that one takes to show it good to do—to please a friend, to win a prize, to pass the time. It is no accident that these voluntary states of mind do not represent the subject's take on their object.

<sup>22</sup> This will be the same class Scanlon calls “judgment-sensitive attitudes.” See his *What We Owe*, especially the first chapter. I argue for a slightly different account (and a different label) in “The Will as Reason” (ms).

not voluntary, it would be a crude mistake to think that they are therefore *involuntary*, in the way that sneezing or blinking is involuntary, or in the way that one's digestion is involuntary.<sup>23</sup> They are not states that simply befall one, or processes to which one is passive. This is perhaps most obvious if we consider intention.<sup>24</sup>

Intending, one might be tempted to think, is surely different than believing, fearing, resenting, being angry, or being jealous. While the latter class of attitudes seems to involve some captivity to the world or to their object, intending seems wholly at the discretion of the intending subject. One might then think that intending is voluntary—something we can decide to do for any reason we take to show it good to intend.

This is a tempting mistake. Intending fails to be voluntary in just the same way believing, resenting, and fearing fail to be voluntary. An intention, no less than a belief, expresses a rational subject's take on its object—in this case, on some possible action. To intend to  $\phi$  is to have settled for oneself the question of whether to  $\phi$ . But one cannot settle that question for reasons which one takes to bear only on the distinct question of whether *intending* is in some way good to do.

Intending can seem more voluntary not only because it is always easy to confuse an attitude with its object (in this case, voluntary action), but also because the question of whether to perform some action or bring about some state of affairs—whether to take a walk or end the fight—is an extremely permissive one. A subject can settle such a question for any reason the subject takes to show the action good to do, including the fact that intending to perform that action would be in some way good. If you are upset

---

<sup>23</sup> Of course, it would also be a mistake to think that they fail to be voluntary in the way that actions done under duress or in ignorance fail to be voluntary.

<sup>24</sup> Again, I make the points below in “Controlling Attitudes” and “The Wrong Kind of Reason.”

because I have no intention to attend your party, that very fact can be reason enough to attend. That is, the fact that the intention would please you, a fact that shows *intending* good to do, can be a reason to attend the party—it can show *attending* good to do.<sup>25</sup>

But in an important range of cases, the fact that intending to  $\phi$  would be good is not taken to be a reason to  $\phi$ . In such cases, it becomes clear that intending is on a par with believing and other commitment-constituted attitudes. One such case is provided by Gregory Kavka's well-known "toxin puzzle,"<sup>26</sup> but such cases occur in less unusual circumstances. Perhaps I am unhappy that you do not intend to marry me—and perhaps I care less about the social and legal arrangement than about your state of mind, the fact that you have no such intention. And perhaps you would be willing to adopt a state of mind that pleases me—perhaps you would be willing to house a marriage commitment, so to speak—so long as you do not need to actually engage in the matrimonial and enter the legal condition. It would seem that you could not, under these conditions, intend. The reasons you have for intending are not, by your own reckoning, reasons for acting. But intending just is committing to an action.

So an intention, no less than a belief, is a kind of commitment to its object—a commitment to acting in some way. But one cannot commit to acting for reasons that one does not, oneself, take to show the action worth doing, any more than one can commit to  $p$  as true for reasons that one does not take to show  $p$  true. So, upon examination, it becomes clear that intention belongs in the class with belief and

---

<sup>25</sup> If one doubts, here, that the concern is really with the intention, we can rig a science-fictional case to make it so: you have an intention detector, and will reward me handsomely for intending to drink some water. I can win the reward, by deciding to drink the water. The fact that the intention is good to have has made the action worth doing.

<sup>26</sup> Gregory Kavka, "The Toxin Puzzle," *Analysis* 43 (1983): 33–36.

resentment—it is an attitude that a thinking subject takes towards some object. Precisely because it represents the subject’s take on the object, it cannot be adopted by deciding to adopt it, where one could decide to adopt it for any reason one took to show adopting it good to do.

If intending is not voluntary in just the same way that believing is not voluntary, the scope of the voluntary should start to seem considerably narrower than the scope of the active or the range of things for which we are responsible. Again, to say that commitment-constituted attitudes are not voluntary is not to say that they are involuntary, or something in relation to which we are passive. While it might at first seem that believing, fearing, resenting, and the like are somehow captive to their object, while intending is at the discretion of the subject, in fact all these attitudes equally represent the subject’s take on their object; the differences between them simply reflect the different “takes” that a subject can have on an object—the subject can decide whether it is true, or fearful, or worthy of resentment, or whether to do it. The thinking subject can settle any of these questions however he or she sees fit, and so arrive at, maintain, or revise these attitudes however he or she will.<sup>27</sup>

The only thing we have seen the thinking subject unable to do is settle a question for reasons that he or she does not take to bear on it. Rather, by finding certain reasons convincing, the subject will always be understood to settle the question on which he or she takes *those* reasons to bear.<sup>28</sup> If one finds convincing the reasons that one takes to

---

<sup>27</sup> We cannot, of course, form or revise them in ways that we do not see fit. Though sometimes one hears complaints about this lack of “freedom,” they are unwarranted.

<sup>28</sup> A background assumption is that a reason is a consideration that bears on a question. Thus reasons are always already associated with a particular question. By finding a reason (or set of reasons) convincing, the subject will settle *that* question. I argue for this conception in “The Wrong Kind of Reason.”

bear on whether *p*, one will, therein, settle the question whether *p*. If instead one finds convincing reasons that one takes to show it worthwhile or in some way good to settle positively the question of whether *p*, one will, for *those* reasons, therein settle the question on which *they* are taken to bear, viz., the second-order question of whether it is in some way good or worthwhile to settle positively the question of whether *p*. In suitable circumstances, one might then also decide to bring it about that one settles positively the question of whether *p*.

Because one can only settle a question for reasons one takes to bear on it, attitudes that embody one's answer to some (range of) question(s) can only be formed or revised for reasons one takes to bear on the relevant question(s). But this is no limitation on agency; it is simply a fact about how we understand what someone has done, in finding certain reasons convincing. The apparent limitation on our agency—the fact that we cannot believe whatever we think it good to believe, or intend whatever we think it good to intend, or fear for reasons that show fear useful—only reflects the fact that these attitudes are understood as our answers to questions *other than* the question of whether it is good to have that very attitude.

The fact that we cannot believe whatever we think it good to believe, intend whatever we think it good to intend, fear for reasons we only take to show fear useful, etc., seems to be a limitation on our agency only because we are accustomed to thinking about the sort of agency we exercise over our own actions, over ordinary objects, and over external states of affairs. We are accustomed to thinking that we control a thing—whether our right hand or our furniture—through some sort of voluntary activity. If we think it good to raise our right hand, we can decide to do that, and, providing all goes well, do it. If we



think it good to rearrange the room, we can decide to do that, and, providing all goes well, do it. As soon as we are able to think about our own mind, we can think about what would be good with respect to *it*. We can find ourselves with reasons that show it good to believe *p*, intend to  $\phi$ , or resent *x*, which we do not take to show that *p*, to show  $\phi$ -ing worthwhile, or to show *x* worthy of resentment. We then think we should be able to control our own mind, in response to these reasons. If we think it good to believe *p*, we think we should be able to decide to believe *p*, and so believe it.

As noted, we *can* do this, in one sense: we can exercise the same sort of control over our minds that we exercise over ordinary objects, like the furniture. We can act in ways designed to produce this or that change in our mind, to bring about this or that belief or intention or other commitment-constituted attitude. Our ability to manage and manipulate our own mind is limited only by our cleverness, luck, and industry—only by our ability to design ways to successfully bring it about that we settle the relevant questions in the required ways.

But typically we are not satisfied with this merely manipulative, managerial control over our own minds. We think we should be able to believe, or intend, or fear, or resent, in the way we act—not merely in the way we manipulate external objects. We think we should be able to believe, or intend, or fear, or resent, by deciding to, for practical reasons. After all, we sometimes do control our minds in this way: we can often remember where we left the car, or imagine world peace, or bisect a triangle in our imagination, by deciding to, for practical reasons. The fact that we cannot believe that *p*, by deciding to, for practical reasons, in the way we can raise our right hand (or rearrange the furniture or imagine that it is raining) by deciding to, for practical reasons, can strike

us as an affront to our abilities; we think we have encountered a shortcoming in our powers. I hope to have gone some distance towards debunking this thought (a thought likely to tempt any creature able both to think about its own mind and to act). We cannot control our beliefs, intentions, or other commitment-constituted attitudes in the way we control our actions, not because of any shortcoming in our agency or limitation in our powers, but rather because each of these attitudes embodies our answer to some question other than whether that very attitude is in some way good to have or adopt. But we can answer the relevant question(s) however we see fit.

Because these attitudes embody our answer to some question(s), we exercise a distinctive form of control over them. While we do not control them in the way we control our actions, they are not, for that states to which we are passive, or simply things which we can only affect and manage through our actions, like our furniture or our allergies. Because these attitudes embody our take on the world, on what is or is not true or important or worthwhile in it, we control them by thinking about the world, about what is or is not true or important or worthwhile in it. Because our minds change as our take on the world changes—because our minds change as we change our minds—we can be said to be “in control” of our commitment-constituted attitudes.<sup>29</sup> This is an odd kind of control, to be sure: the object controlled and the subject who controls are not distinct. There is no reflective distance in the offing. We change our minds, and so control our attitudes, not by reflecting on or thinking about our mind, but rather by thinking about the

---

<sup>29</sup> A similar point has been emphasized by Richard Moran (though he argues against thinking of this as a form of “control”). See his *Authority and Estrangement: An Essay on Self-Knowledge* (Princeton: Princeton University Press, 2001), especially 114–20.

object of our thoughts. The controlling happens “behind the lens,” so to speak. The thinking subject controls its thoughts in thinking them.<sup>30</sup>

#### RESPONSIBILITY AND CONTROL

One’s commitment-constituted attitudes, then, embody one’s take on things, on whether some object is true, good, dangerous, offensive, worthwhile, rightly one’s own, to be brought about, etc. But, precisely because they embody one’s take on an object, they are not, and (I hope I have shown) could not be, voluntary—they are not the sort of thing one can form or revise or maintain for any reason one takes to show it worth doing. And yet, for that same reason—because they embody one’s take on an object, one’s answer to some range of questions about what is true, good, worthwhile, important, dangerous, wrong, to be done—these attitudes comprise that stretch of one’s mind that is the object of epistemic and moral assessment and reaction. Thus, if being responsible is a matter of being both open to such assessments and (depending on the outcome of that assessment) the appropriate target of certain reactions, then the class of things for which one is most fundamentally responsible—the class of things that can constitute the object of that assessment and reaction—are also necessarily non-voluntary.<sup>31</sup>

This may seem a puzzling conclusion. We started with an apparent tension between the thought that one is responsible for one’s beliefs and the thought that belief is not voluntary. The underlying assumption, generating the tension, was that one can be

---

<sup>30</sup> The possibility of reflective distance has widely been thought essential for moral responsibility. While this may be so, I would argue it is not because the kind of control we exercise over the objects of moral assessment involves reflection. (I hope to write a paper on this matter in the near future.)

<sup>31</sup> I have, in effect, argued that one’s will could not be in one’s voluntary control, given the older notion of the will and the more modern notion of voluntary. (See note 11.)

responsible only for what voluntary. We have now arrived at a view—I hope a plausible view—according to which that for which one is most fundamentally responsible could not be voluntary. I will now try to dispel the air of paradox by suggesting how the requirement of voluntariness gains its plausibility: voluntariness often plays an important role in one’s responsibility for things that are neither commitment-constituted attitudes nor their immediate, non-voluntary expression—for what I will call *external conduct* and *external states of affairs*—because, unless these things are in one’s voluntary control it is hard to see how they could accurately reflect the quality of one’s will. If some stretch of external conduct—putting the letter in the mail, standing on the hose, stepping down from office—is not something you have done voluntarily—that is, by (in some sense) settling the question of whether to do it—then it is hard to see how that bit of conduct could display your moral personality. And so it is hard to see how you could be a fit object of moral assessment or of the reactive attitudes on account of it. If some external state of affairs—the horrible mess in the kitchen—was neither something that you intentionally brought about nor something that was the foreseen outcome of something you intentionally did, then it is hard to see how its coming about reflects the quality of your will.<sup>32</sup> Likewise, if we hold you responsible for the perdurance of the mess—if its continued existence is taken to reflect something about your moral personality—it seems plausible to say that eliminating the mess must be something you can do by deciding to expend effort to that end—something you can do voluntarily. Otherwise it is hard to see

---

<sup>32</sup> I say it is hard, but not impossible. Both non-voluntary external conduct and states of affairs that are not the intended or foreseen outcome of voluntary action can reflect the quality of your will if it was conduct or a state of affairs with respect to which you had some obligations. If you neglect these obligations, then that fact will show something about the general quality of your will, or your moral personality. Further, it seems likely that we can have obligations only for actions that can be voluntarily performed. I will not take up these further interesting issues here.

how its existence could reflect the quality of your will. So it can seem quite plausible to say, of things other than commitment-constituted attitudes and their immediate non-voluntary expression, that they must be in your voluntary control before you are responsible for them—because such control secures a connection between a stretch of external conduct or an external state of affairs and one’s moral personality.<sup>33</sup>

It would be a mistake, however, to extend this thought to commitment-constituted attitudes. These attitudes need not be voluntary in order to reveal one’s moral personality. Rather, they constitute one’s moral personality—moreover, they do so precisely because they represent one’s take on the world, and so could not be voluntary. The fact that we do not exercise voluntary control over these attitudes need not entail that we do not exercise any control over them. One must simply allow that there are other forms of control.<sup>34</sup> These attitudes represent one’s take on what is true or important or worth doing; if one changes one’s take on those things, one’s mind therein changes. We could say, then, that this stretch of one’s mind is in one’s control because it is constituted by the conclusions one comes to.

#### CONCLUSION

We started with an apparent tension: beliefs are not voluntary and yet are aspects of ourselves for which we are responsible, on account of which we are open to various forms of assessment and response. The apparent tension, I suggested, is predicated on a

---

<sup>33</sup> Scanlon makes a similar point when he says that certain conditions render “moral appraisal inappropriate because they break the usual connection between the action and the judgment-sensitive attitudes of the agent,” or, in our terms, between the external conduct and the moral personality. See *What We Owe*, 277.

<sup>34</sup> David Owens considers this strategy and argues against one implementation of it (what he calls “reflective control”) in *Reason without Freedom: The Problem of Epistemic Normativity* (London: Routledge, 2000). He concludes that we do not need to be free in order to be responsible. I would hope to argue for a different implementation of the strategy.

false assumption, namely, that we are responsible only for what is voluntary. I have sketched an account of what it is to be responsible and of what it is for a thing to be voluntary according to which those things for which one is most fundamentally responsible could not be voluntary. One is most fundamentally responsible for those attitudes that embody one's take on the world—on what is true, good, worthwhile, important, fearful, hateful, delightful, etc. But precisely because these attitudes embody one's take on their object, they could not be under one's voluntary control—they could not be the sort of thing one could do by deciding to do it. Beliefs, then, are a central example of the sort of thing for which we are most fundamentally responsible, in part because they are the sort of thing which could not be voluntary.<sup>35</sup>

---

<sup>35</sup> This paper has benefited from the help of many, including Jason Baehr, Sarah Buss, Stephen Finlay, Mark Greenberg, Barbara Herman, Aaron James, Sean Kelsey, Adam Leite, Herbert Morris, Nikolaj Nottelmann, Angela M. Smith, my graduate seminar at UCLA, and the Southern California Law and Philosophy Discussion Group.