Matthias Steup  
Justification, Deontology, and Voluntary Control

In contemporary epistemology, it is widely rejected that epistemic justification—the kind of justification traditionally considered a necessary condition of knowledge—is by nature deontological.¹ I shall present a brief outline of the deontological account of justification (epistemic deontology, for short) and then discuss the two arguments on which its rejection is based.

Deontological Justification Defined

According to epistemic deontology, the concept of justification is to be defined in terms that belong to the family of deontic locutions such as obligation, prohibition, or permission, each of which can in turn be defined in terms of an ought. To say that ϕ-ing is obligatory means that one ought to ϕ, to say that ϕ-ing is prohibited means that one ought not ϕ, and to say that ϕ-ing is permitted means that it’s not the case that one ought not ϕ. Taking a justified belief to be a permissible belief, let us define justification as follows:

\[ J_d 1 \quad S \text{ is justified in believing that } p =_{df} \text{ It is not the case that } S \text{ ought to refrain from believing that } p. \]

Compare \( J_d 1 \) with an alternative definition, suggested by William Alston’s claim that deontological justification is most centrally freedom from blameworthiness:²

\[ J_d 2 \quad S \text{ is justified in believing that } p =_{df} \text{ It is not the case that, in believing that } p, S \text{ is blameworthy.} \]

Whether \( J_d 1 \) and \( J_d 2 \) are equivalent depends on how we understand the notion of blameworthiness. Let’s distinguish between a weak and a strong sense of the notion:

\[ B_w \quad S \text{ is blameworthy for } \psi \text{-ing} =_{df} S \text{ ought not } \psi. \]

¹ Trusting the reader will not need to be constantly reminded of what kind of justification this paper is about, I will henceforth omit the qualifier ‘epistemic’, except for contexts in which adding it is called for to distinguish epistemic from other kinds of justification.
² An alternative would be to take justified belief to be obligatory belief. But this does not adequately capture the meaning of the word ‘justification’. William Alston puts the point aptly when he says the following about the justification of an action: “To say the action was justified does not imply that it was required or obligatory, only that its negation was not required or obligatory. This holds true whether we are thinking of moral, legal, institutional, or prudential justification of actions.” Alston adds: “The most natural way of construing the justification of beliefs is in parallel fashion.” Alston 1989, p 125f.
³ In his classic “Concepts of Epistemic justification, Alston says “\( J_d \) is, most centrally, a concept of freedom from blameworthiness, a concept of being ‘in the clear’ so far as one’s intellectual obligations are concerned.” Alston 1989, p. 89.
S is blameworthy for ϕ-ing =df S does not deserve to be blamed for ϕ-ing.

If the word ‘blameworthy’ in J₂ is understood in the sense of B_w, J₂ is equivalent to J₁; if the word is understood in the sense of B_s, J₂ is not equivalent to J₁. In fact, in that case, J₂ yields an account of justified belief that should be rejected.

Suppose we understand being justified in ϕ-ing as not deserving to be blamed for ϕ-ing. This understanding distorts what the concept of justification ordinarily means. The distortion is due to the fact that when one’s ϕ-ing constitutes a rule infraction but one has an excuse for ϕ-ing, then one does not deserve to be blamed for ϕ-ing. Hence, if we define being justified as freedom from blameworthiness in the strong sense, we get the following result: whenever an agent has an excuse for ϕ-ing, the agent’s ϕ-ing is justified. This is a bad result since having an excuse is not a condition that gives one justification for ϕ-ing.

Suppose Sam didn’t pay his rent on time because at the end of a vacation a volcano eruption delayed his flight back home by a week. The unforeseeable delay gives Sam an excuse, but not a justification, for not paying his rent on time. The sort of thing giving him justification would be, for example, his landlord’s permission to pay his rent on the 15th instead of the 1st. The same point applies to beliefs. A severe headache, sleep deprivation, or detrimental side-effects of a drug might give us an excuse for believing something silly. Such conditions, however, do not give us justification for the silly belief. Hence, if J₂ is construed in terms of B_s, the result will be a defective account of justified belief.

Note that there is no inconsistency in classifying an action or a belief as blameworthy in the weak sense while not being blameworthy in the strong sense. Sam’s not paying his rent on time was unjustified or blameworthy in the sense that Sam didn’t do what he ought to have done, namely pay his rent on time. But since he had an excuse, it would be unfair to blame him for this omission. Likewise, if due to the side-effects of a drug you believe something silly, you hold an unjustified or blameworthy belief in the sense that you believe something that you ought not believe. Saying this is entirely compatible with the concession that, since you have an excuse, we shouldn’t blame you for this belief.

To sum up: If we understand J₂ in terms of B_w, then J₁ and J₂ are equivalent. In that case, nothing substantial rides on which definition we employ. If, however, we understand J₂ in the sense of B_s, the result will be an ill-conceived notion of
justification. Deontological justification, then, should not be defined in terms of the strong sense of the concept of blameworthiness.

**Epistemic and Non-Epistemic Justification**

Next, we need to consider how the *epistemic* ought differs from the ought of prudence and morality. The definiens in J₄.₁ should not be taken to mean: “It is not the case that S *prudentially* ought to refrain from believing that p.” Nor should it be taken to mean: “It is not the case that S *morally* ought to refrain from believing that p.” Rather, the definiens of D₄.₁ must be understood this way: “It is not the case that S *epistemically* ought to refrain from believing that p.” The *epistemic* ought differs from the *ought* of prudence and the *ought* of morality, and thus the definiens of D₄.₁ implies nothing with regard to the question of what S ought to believe prudentially or morally. Exactly how, though, does the epistemic ought differ from the ought of prudence or morality?

Let us approach the question this way: Advocates of epistemic deontology would say that what one epistemically ought or ought not believe is a function, not of one’s moral or prudential reasons, but of one’s epistemic reasons. How, though, do epistemic reasons differ from moral or prudential reasons? The standard approach to answering this question is to invoke the truth goal. Exactly how the truth goal is to be invoked is not an easy matter to settle. Bracketing difficult issues that would lead us astray in the present context, I will just say the following: Epistemic reasons are the kind of reasons that are relevant when we aim at the goal of believing a proposition p only if p is true, and thus are the kind of reasons that confer probability or likelihood of truth on our beliefs. In this regard, epistemic reasons differ from moral or prudential reasons, for neither of the latter kind of reasons confers probability on our beliefs, or makes our beliefs likely.

**Why Favor Deontological Justification?**

Not all normativity is deontological. We can, for example, evaluate agents with regard to their physical suitability for certain athletic activities. We might say that with regard to height and weight, Al is suited better than Ben for playing basket ball,

---

4 Since one’s reason are one’s evidence, epistemic deontology thus understood is a version of evidentialism, according to which whether what one believes is justified is determined by one, and only one, thing: one’s evidence. Cf. Conee and Feldman 2004.

5 For an illuminating discussion of the relevant issues, see David 2001.
whereas Ben is better suited than Al for the sport of wrestling. Making such judgments does not entail anything deontological. It does not entail that Al had an obligation to acquire fitness for the game of basketball, or that it was permissible for Ben to have physical characteristic making him suitable for the game of wrestling.

Likewise, some epistemologists favor evaluating an agent’s performance as a believer in a fashion that is analogous to the way we evaluate an agent’s suitability for a certain kind of sport. According to this non-deontological approach, we might want to evaluate one belief as justified and another as unjustified because we think that the first, but not the second, serves the pursuit of truth, without intending this evaluation to imply that one is permitted or even obliged to hold the first but not the second belief. What motivates such a non-deontological approach to justification are, primarily, two arguments:

*The Argument from Lacking Truth Conduciveness*

Deontological justification is epistemologically irrelevant because it is not truth-conducive.\(^6\)

*The Argument from Doxastic Involuntarism*

Beliefs are unsuitable for deontological evaluation because we lack voluntary control over them.\(^8\)

I will discuss these arguments in due course. For now, suppose deontologists have the resources for an effective rebuttal of them. This rebuttal, by itself, does not explain why one might, to begin with, favor the deontological over the non-deontological approach. What, then, is the primary motivation recommending the deontological understanding for justification?

In a nutshell, the answer is that deontologists view belief formation as a form of agency. A person’s physical characteristics, such as her height, hair color, or eye color fall outside the scope of agency. From the deontological point of view, beliefs do not. Believing is a form of agency, and thus we bear responsibility for our beliefs just as much as we bear responsibility for our actions. Just as there are things we ought to do and ought not do, there are propositions we ought to believe and others we ought not believe.

A secondary motivation for endorsing a deontological understanding of epistemic justification arises from the fact that deontological evaluation of our

\(^6\) Cf. Alston 1989, p. 96f.
\(^7\) Ibid, pp. 95 and 145f.
\(^8\) Ibid, p. 95, 119-136.
doxastic conduct is deeply entrenched in our ordinary linguistic practices. When someone states a belief we consider irrational, we might say ‘You shouldn’t believe this.’ Similarly, when someone refuses to believe what we consider being supported by excellent evidence, we might easily say ‘You ought to believe it.’ There would be nothing linguistically infelicitous about saying such things. If those who reject epistemic deontology were right, then we would need an error theory: a theory explaining how it can be that, in our ordinary practices of evaluating each other’s performance as believers, we employ a framework that is fundamentally mistaken inasmuch as it takes belief to be something it isn’t, namely a form of agency. Advocates of epistemic deontology enjoy the theoretical advantage of not having to supply an error theory of that kind.

The Argument From Lacking Truth-Conduciveness

According to the first of the two primary anti-deontological arguments mentioned above, deontological justification is not truth-conducive and thus fails to be epistemic justification, the kind of justification that is relevant in epistemology. To establish this point, Alston describes two situations in which a subject enjoys deontological justification for a belief without having reasons in support of the belief that confer likelihood on it. I will discuss only the first of Alston’s two examples, since what I will say about it can easily be applied to the second as well. Alston’s first example presents us with a culturally isolated tribe. Suppose members of this tribe base a belief about the outcome of a battle on reading the entrails of a dead animal. What these entrails reveal, according to the tribe’s experts on predicting the future in this fashion, is that they will win the battle. Let’s use ‘W’ to refer to this belief. Alston’s argument can be summed up as follows:

(1) Deontological justification is essentially freedom from blameworthiness.
(2) In believing W, the tribe members are free from blameworthiness.
   (Since they are culturally isolated and thus have never been exposed to criticism of their cognitive practices, they do not deserve to be blamed for believing W.)

Therefore:

(3) The tribe members have deontological justification for believing W.

Alston 1989, p. 95.
Epistemic justification for believing W requires possessing a reason that confers likelihood on W.

The tribe members do not possess a reason that confers likelihood on W. (Their only reason for W is a proposition about what the dead animal’s entrails say, and that reason does not confer likelihood on believing W.)

Therefore:

The deontological justification the tribe members have for believing W is not epistemic justification.

Above, I argued that we must distinguish between a strong and a weak sense of ‘blameworthiness’:

\[ \text{B}_w \quad \text{S is blameworthy for doing } x \iff \text{S ought not do } x. \]

\[ \text{B}_s \quad \text{S is blameworthy for doing } x \iff \text{S ought to be blamed for doing } x. \]

If we understand ‘blameworthiness’ in the strong sense, then the first premise is false. If we understand ‘blameworthiness’ in the weak sense, then the second premise is false.

Suppose we understand the first premise in the strong sense of ‘blameworthiness’. As I argued above, we then get the implausible consequence that, whenever a subject has an excuse for a silly belief, her silly belief is justified. But what justifies a belief is one’s evidence, not the kind of conditions that give one an excuse for believing something silly. Hence a B\(_s\) construal of first premise must be rejected.

Now suppose we understand the first premise in the weak sense of blameworthiness. In that case, we should judge that the tribe members’ belief is blameworthy, a belief they ought not hold, for the belief is based on not probability-conferring reasons but the unreliable method of interpreting the entrails of a dead animal. In making this judgment, we need not also judge that the tribe members ought to be blamed for believing W. Rather, even though we think they ought not believe W and are blameworthy for believing W in that sense, we could easily agree with Alston that, due to their long-standing customs and traditions, firmly in place due to cultural isolation, the people of this tribe ought not to be blamed for believing W. There is no inconsistency involved at all in making these two

---

\(10\) Externalists would of course say that it’s not one’s evidence but some external condition such as reliability that justifies one’s beliefs. But, I take it, they would agree that conditions providing an excuse are not among the conditions that justify our beliefs.
judgments.\textsuperscript{11} Hence, given a B\textsubscript{w} reading of the first premise, the second premise of Alston’s argument must be rejected.

**Doxastic Involuntarism and the Datum Supporting It**

Next, I turn to an examination of the other primary objection to epistemic deontology, the Argument from Doxastic Involuntarism. Doxastic involuntarism is the view that beliefs—or more generally, our doxastic attitudes such as belief, disbelief, and suspension of judgment—differ from actions in the following respect: Whereas we have voluntary control over what we do, we lack voluntary control over what we believe. According to the Argument from Doxastic Involuntarism, introduced to the epistemological literature by William Alston, the significance of doxastic involuntarism lies in its incompatibility with epistemic deontology. If indeed beliefs differ from actions in being beyond our control, then belief is not a form of agency and thus unsuitable for deontological evaluation. More formally, the argument may be stated as follows:

1. If our doxastic attitudes are suitable for deontological evaluation, then we have voluntary control over them.
2. We do not have voluntary control over our doxastic attitudes.
   
   Therefore:
3. Our doxastic attitudes are unsuitable for deontological evaluation.

Sharon Ryan rejects both premises of this argument.\textsuperscript{12} The first premise raises certainly interesting and difficult issues. Since an evaluation of it is beyond the scope of this paper, I will focus just on the second premise only. I will begin by reviewing why this premise enjoys a good deal of initial plausibility and then proceed to argue that there are good reasons to reject it.

Alston’s defense of the second premise is based on examples: You cannot now believe that the United States is still a colony of Great Britain.\textsuperscript{13} When you see and feel that it is raining, you cannot refrain from believing that it is raining.\textsuperscript{14} Such examples generalize to what I will call the ‘datum’ in support of doxastic

---

\textsuperscript{11} About an earlier response of mine to his cultural isolation case (Steup 1988), Alston says “I think that Steup is displaying an insensitivity to cultural differences.” Alston 1989, p. 146. The response offered here, it seems to me, is not a proper target for the charge of cultural insensitivity.

\textsuperscript{12} Ryan 2003.

\textsuperscript{13} See Alston 1989, p. 122.

\textsuperscript{14} See ibid, p. 129.
involuntarism: You cannot believe what in light of your evidence is clearly false, and you cannot refrain from believing what in light of your evidence identifies clearly true. In addition, there is a third category: propositions that, given your evidence, are neither clearly false nor clearly true. Propositions of this kind are such that you can neither believe nor disbelieve them. I will use the following three propositions to illustrate the datum to which Alston appeals:

\[
\begin{align*}
H & \quad \text{I have hands.} \\
W & \quad \text{I have wings.} \\
D & \quad \text{The number of ducks is even.}
\end{align*}
\]

Given my evidence, H is clearly true and W is clearly false. Hence I cannot believe W and cannot refrain from believing H. D, in light of my evidence, is neither clearly true nor clearly false. Hence I can neither believe D nor disbelieve D, which is to say that I cannot refrain from suspending judgment about D. The vast majority of our doxastic attitudes falls into one of these three categories and hence would appear to be beyond our control. In a nutshell, then, Alston’s datum can be expressed thus:

Given our evidence, we cannot believe otherwise. In light of this datum, it would seem that the second premise is true.

The Two Ingredients of Voluntary Control

To assess whether the datum really succeeds in establishing the truth of doxastic involuntarism, we must take note of the fact that voluntary control involves two aspects: executional and volitional control. Suppose you are in an auditorium listening to a lecture. If you have control over your whereabouts, then it must be the case that, if you choose to stay, you will not be forcibly removed from the premises, and if you choose to leave, you will not be prevented from doing so by locked doors. We may define this kind of control as follows:

Executional Control

One has voluntary control over \( \varphi \)-ing \( \rightarrow \) If one decides to do \( \varphi \), one can \( \varphi \), and if one decides to refrain from \( \varphi \)-ing, one can refrain from \( \varphi \)-ing.

If you can no longer stand listening to an excruciatingly boring lecture and decide to leave, then being tied with ropes to your seat, experiencing sudden paralysis of your legs, or encountering locked doors stand in the way of executing your decision, thus robbing you of executional control over whether to leave or not. Similarly, if the lecture is fascinating and you want to continue listening to it, but security personal
forcibly removes you from the auditorium, then you cannot execute your choice to remain in the lecture hall. In that case that you no longer enjoy executional control over whether to stay or not.

Here are two further examples of executional control failure:

**Paralyzed Peter**

Due to a stroke, Peter encounters sudden paralysis of his right arm. He wants to raise it to reach for his coffee, but his arm won’t move.

**Twitching Tom**

His eye lid twitches. Tom is trying hard to stop his eye lid from twitching, but with no success.

Peter and Tom are suffering from executional control failure. Hence Peter doesn’t have voluntary control over the behavior of his arm, and Twitching Tom does not have voluntary control over the behavior of his eye lid. Executional control failure, however, is not the only way in which one can be prevented from enjoying voluntary control. Consider two further examples:

**Agoraphobic Al**

His friends are taking a walk in the park, but Al can’t join them because he suffers from agoraphobia, an excessive and irrational fear of wide, open places. Due to his mental illness, he cannot decide to do an unnerving thing as going to the park.

**Mysophobic Mel**

Although Mel has perfectly clean hands, he just washed and then disinfected them for the 67th time since he got up in the morning. He did this because he suffers from mysophobia, an excessive and irrational fear of germs and contamination. Due to his mental illness, he decides, up to 100 times a day, to wash and disinfect his hands.

The problem Al and Mel encounter is not that of being unable to do what one has decided to do, that is, that of being unable to execute one’s decision. It’s not the case that Al doesn’t take a walk in the park because each time he tries to do that muscle spasm prevent him for doing it, or civil unrest on the streets would make leaving one’s house unwise. Rather, Al’s problem is that, due to his mental illness, he cannot decide to take a walk in the park. Similarly, the reason why Mel is washing his hands is not that he is forced at gunpoint to do so, or that a series of muscular convulsions miraculously results in an episode of undesired hand washing.
Rather, due to his mental illness, Mel cannot help but decide to wash his hands once again. Al and Mel, then, suffer from a loss of volitional control—control over one’s decisions and choices—which we may defined as follows:

**Volitional Control**

One has voluntary control over \( \varphi \)-ing \( \rightarrow \) One can decide to do \( \varphi \) and one can decide to refrain from \( \varphi \)-ing.

To sum up, voluntary control breaks down into two parts: the ability to execute one will, and control over one’s will itself. Each of them is a necessary condition of voluntary control.

**Why We Can’t Believe Otherwise**

Let us return to Alston’s datum, our inability to believe otherwise given the evidence we have. Is this inability due to control failure of the executional or the volitional kind? Let us consider my inability to believe that I have wings.

Given my evidence—I see and feel that there are no wings that are attached to my body—I am forced to disbelieve the proposition that I have wings. Now, am I forced to disbelieve this proposition

(a) because I cannot decide to believe a crazy thing like that

or

(b) because, upon deciding to believe such a crazy thing, I cannot execute my decision?

Here are four reasons for thinking that (a) is correct. First, in our personal experience, typically we are not familiar with episodes in which we resolved to believe or disbelieve something and then find ourselves unable to execute that choice. Although such cases are not impossible, they are quite rare.\(^\text{15}\)

Second, under normal circumstances, it is easy to execute one’s doxastic choices. It’s even easier than turning the light on or off. On the way to the light switch, one might fall and brake one’s neck. Other people might try to prevent one from flipping the switch. No such risks are involved in executing a doxastic decision.

The kind of thing that might block a doxastic volition to believe \( p \) are an emotional

\(^{15}\) Consider a subject who, having been brought up in a devout family, takes at college a course in the Philosophy of Religion and, upon studying the problem of evil, decides that the God he used to believe in does not exist. In a case like this, it is easily imaginable that the theistic belief ‘sticks’, that it does not easily go even though the subject is trying to get rid of it.
aversion to \( p \) or a deep commitment involving the rejection of \( p \). However, executional control failure due to such phenomena is rare.

Third, supposing I could somehow decide to believe I have wings, why should my evidence, having thus far pitifully failed to prevent me from being irrational, finally spring into action and block my endeavor when it comes to putting the crazy belief into my belief box? This would be a rather strange and mystifying explanation of my inability to believe that I have wings.

Fourth, doxastic and practical irrationality are similar. Suppose that, listening to a philosophy talk in which the speaker advocates a view which you strongly disagree, you consider taking off a shoe and throwing it at the speaker. Suppose further that you couldn’t do an impolite and irrational thing like that. What explains your inability? Would it be

(a) because you can’t decide to do it

or

(b) because you are unable to execute the decision to do it? (We might imagine that, knowing your deed will meet with public disapproval, you get so nervous that you find yourself unable to take off a shoe.)

Again neglecting rare and unusual cases, it seems obvious that (a) is the correct answer. In general terms, the point is that, if your inability to make a bad practical decision explains why you can’t perform an obviously irrational act, then we have confirmation for thinking that your inability to make a bad epistemic decision explains why you cannot acquire an obviously irrational belief.

These reasons apply with equal force to the other examples illustrating Alston’s datum. We cannot refrain from believing that we have hands, and we can neither believe nor disbelieve that the number of ducks is even because, given our evidence, we cannot make the decisions required for adopting these doxastic attitudes. What explains the datum, then, is that our evidence makes it difficult, if not psychologically impossible, for us to decide in favor of attitudes that clearly conflict with it.\(^{16}\)

\[^{16}\text{H. H. Price appears to express agreement with this point when he writes: “If you are in a reasonable frame of mind, you cannot help preferring the proposition which the evidence favors, much as you may wish you could.” Price 1954, p. 16.}\]
Compatibilism and Incompatibilism about Volitional Control

We now have an improved understanding of Alston’s datum. The reason why we cannot deliberately believe contrary to what our evidence dictates is that we cannot decide to adopt doxastic attitudes that are clearly inconsistent with our evidence. Does this datum support doxastic involuntarism in any straightforward fashion? To discuss this question, we must briefly consider the issue of free will and determinism. Compatibilists and incompatibilists agree that one has voluntary control over ϕ-ing only if one has executional control over ϕ-ing. For example, they would agree that, if I want to leave but the doors are locked, I don’t have control over my whereabouts. What libertarians and compatibilists do disagree about is what it takes to have volitional control. Here is a brief statement of each view:

Incompatibilism
Volitional control requires the ability to decide otherwise under the very same circumstances. This requirement is satisfied only in the absence of causal determination. In a deterministic world, control over our decisions and choices is impossible.

Compatibilism
Control over our decisions and choices is possible even in a deterministic world. Volitional control does not require the ability to decide otherwise under the very same circumstances. What it requires is not the absence of causal determination but rather causal determination of the right kind.

The point of compatibilism is to allow for free decisions, or control over one’s will, in a deterministic world. Compatibilists secure this result by insisting that, when we want to assess whether an agent’s decision was under her control or not, what is relevant is not whether the decision was caused, but rather how it was caused.17 What matters is whether it was caused in a way that enables or prevents control. Let us say that a control-enabling causal history is a good causal history, whereas a control-preventing one is a bad causal history. Obviously, a lot of work needs to go into developing the details of an account of how to discriminate between good and bad causal histories. Here, there is no need to review the array of options available for developing these details. Instead, I will consider just one prominent approach.

---

17 For a useful introduction to the compatibilist literature, see McKenna 2009.
according to which the difference between a good and a bad causal history is constituted by the presence of absence of reason responsiveness.18

The basic idea of this reason responsiveness approach can be illustrated as follows. Suppose you decide to wash your hands. Suppose further the following is true of you: (i) you made this decision because your hands are dirty; (ii) you would also have decided to wash your hands if you were about to prepare a meal in your kitchen, or if you just dispensed an insecticide to deal with an ant problem in your house; (iii) you would not have decided to wash your hands if since the last time you washed them nothing happened to make them dirty or contaminate them with contagious agents. In that case, your decision exhibits reason responsiveness, which entitles us to judge that it was under your control. Mysophobic Mel, in contrast, doesn’t satisfy the requirement of reason responsiveness. His recurring decisions to wash his hands, being rooted in mental illness, do not display the kind of sensitivity to one’s reasons that voluntary control requires. Mel’s hand washing decisions, therefore, are not under his control.

Compare Mel’s decision to wash his hands with your decision to wash your hands. Your decision is no less causally determined than Mel’s. In the circumstances prevailing at the moment the decision is made, neither you nor Mel could have decided otherwise. That, according to reason responsiveness compatibilism, is irrelevant to the question of whether the decisions are voluntary. What matters is that, unlike Mel’s, your decision in favor of hand washing instantiates responsiveness to the relevant range of reasons and thus counts as voluntary.

Generalizing, we may say that, according to the reason-responsiveness approach, voluntary control arises from the sensitivity of an agent’s actions to the agent’s practical reasons.19 Henceforth, I will use the term ‘compatibilism’ to refer to the reason responsiveness version. There are, of course, other versions of compatibilism. The line of reasoning I am going to advance could easily be adjusted to these alternative versions.20

---

18 See Fischer and Ravizza 1998. For alternative approaches, see, for example, Frankfurt 1971 and Strawson 1962.
19 Developing a detailed account of reason responsiveness is a challenging project. One pitfall to avoid is the consequence that irrational conduct—which constitutes some sort of failure to be responsive to good reasons—is invariably involuntary. Surely irrational behavior is not always involuntary. How the reason-responsiveness approach can meet this desideratum is not instantly obvious. See Bayer 2010.
The Compatibilist Case for Doxastic Voluntarism

Suppose you are committed to both compatibilism and doxastic involuntarism. To test the consistency of your view, let’s focus on your belief that you have hands. The question is whether this belief is under your voluntary control. As an advocate of doxastic involuntarism, you would say that it is not. But this answer would appear to be in conflict with your endorsement of compatibilism. Surely your belief that you have hands has a good causal history. It originates in visual and tactile perception, which are excellent sources of cognition. And we can easily see that your belief exhibits reason responsiveness. It is not true of you that you stick compulsively to the belief that you have hands no matter what your reasons are. If you were presented with visual and tactile perceptions indicating the loss of your hands, you would no longer believe that you have hands. Suppose, however, you wish to stick to your dual commitment. You wish to abandon neither compatibilism nor doxastic involuntarism. In that case, you would have to argue, implausibly, that your belief that you have hands has a bad causal history: that, being rooted in mental illness or similarly bad causes, is not the result of a reason-responsive process of belief formation.

Note that Alston’s datum is representative of belief formation in general. The vast majority of our doxastic attitudes are locked into place by our evidence. Given our evidence, it is not possible for us to adopt alternate doxastic attitudes. Hence, if you take Alston’s datum to establish general doxastic involuntariness, and you endorse compatibilism, you need to bite a particularly unbecoming bullet, namely the consequence that the vast majority of our doxastic attitudes have a bad causal history. You would then have to view our doxastic attitudes as the products of processes rooted in bad causes such as mental illness, addiction, manipulation, brain washing, hypnosis, etc. This line of reasoning would be exceedingly implausible, amounting to a *reductio* of your dual commitment.

If compatibilists wish to avoid this *reductio*, they need to abandon doxastic involuntarism. Under normal circumstances, one’s belief that one has hands is an instance of reason responsiveness and thus has a good causal history. If one’s evidence were inconsistent with hand possession, one wouldn’t believe we have hands. Compatibilists, then, should say that our belief that we have hands is a belief...
that’s under our voluntary control. Analogous reasoning applies to the vast majority of our beliefs, having their causal origin in perception, memory, introspection, and reasoning. Such beliefs have a good causal history. Compatibilists should say that they are under our voluntary control.

**The Argument from Intentionality**

There is a move compatibilists can make to avoid the consequence of having to embrace doxastic voluntarism. They can make intentionality a necessary condition of voluntary control and argue that beliefs don’t satisfy this condition. This argument can be stated as follows:

The Argument from Intentionality

1. For any doxastic attitude $\alpha$, if $\alpha$ is under one’s voluntary control, then $\alpha$ is adopted intentionally.
2. Doxastic attitudes are never adopted intentionally.

Therefore:

3. For any doxastic attitude $\alpha$, $\alpha$ is not under one’s voluntary control.$^{22}$

Initially it seems true that one’s $\phi$-ing is under one’s voluntary control only if it is intentional. Consider the behavior of Twitching Tom’s eye lid. Its twitching is not under Tom’s voluntary control and it is not intentional. Compare this with a normal episode of washing one’s hands, which is both under one’s voluntary control and intentional. Generalizing, it would seem that behavior over which one has voluntary control is intentional action, whereas behavior over which one fails to have voluntary control, such sneezing, slipping on ice, facial tics and the like, do not fall under the category of intentional action. The first premise, then, seems plausible.

In light of Alston’s datum, the second premise seems plausible as well. Your belief that you have hands did not come about because you decided and henceforth intended to believe that you have hands. Rather, it would seem the belief is an automatic response to your evidence. This point generalizes. It seems that our beliefs are nearly always a causal response to various inputs, some of them good, others bad. So intentions don’t seem to play a role in belief production.

---

$^{22}$ Bennet 1990, Feldmann 2001, and Nottelmann 2006 all argue that intentionally held belief is impossible. For a discussion of my 2008 response to the argument, see Booth 2009. For an interesting response to the Argument from Intentionality, one that is different from the one I pursue here, see Weatherson 2008, p. 547. For a useful summary of literature on intentionality, see Setiya 2010.
Contrary to initial appearance, the first premise is false. Further below, I will argue that the second premise is false as well. For now, let us focus on premise (1). While it is certainly true that frequently when we act, we are carrying out some intention or other, it is also true that a lot of what we do is done without antecedently formed intentions. Here are some examples: shifting from 2\textsuperscript{nd} to 3\textsuperscript{rd} gear while driving to work, turning the doorknob or pressing the door handle down while opening a door, and unscrewing the cap of your toothpaste before brushing your teeth. Actions like these are trained, automatic responses that we perform without thinking about them. They are part of sequences that don’t involve a prior intention for each step of the sequence. Yet such actions are not involuntary. Gear shifting, doorknob turning and unscrewing of caps are not behaviors over which we lack voluntary control. So the first premise is false.

Two Kinds of Intentionality

In response to my appeal to automatic actions, advocates of doxastic involuntarism could try to recover a kind of intentionality that automatic actions instantiate but doxastic attitudes do not. For example, involuntarists could employ Searle’s distinction between prior intention and intention in action. About the items on my list of automatic actions, Searle would say that, while they are not examples of carrying out prior intentions, they are nevertheless examples of intentionality because each of them has an intention in action. Advocates of involuntarism could argue that, while there is such a thing as intention in action without prior intention, as exhibited by automatic actions, there is no such thing as intentionality in belief. Beliefs, therefore, do not exhibit any kind of intentionality.

Exactly what phenomenon is picked out by the term ‘intention in action’? Searle offers an interesting explanation of the difference between behavior where an intention in action is present and behavior where it is not present. The difference, Searle suggests, lies in the experience of acting. In so-called ‘Penfield Cases,’ an experimenter manipulates a subject’s brain so as to make the subject’s arm go up.

\textsuperscript{23} Searle (1983), pp. 84ff. In his 2008, although as an opponent of doxastic involuntarism he agrees with quite a bit of what I say in my 2000 and 2008, Weatherson argues that, contrary to what I am claiming, automatic actions are intentional after all. He does not, however, distinguish between prior intention and intentions in action. If what he has in mind is the latter, then I would agree with him that automatic actions are intentional in the second, weaker, sense of intentionality.

\textsuperscript{24} Ibid, p. 88f.
When the subject is afterwards asked whether he raised his arm, the typical answer is: “I didn’t do that; you did.” The subject can assert this with complete assurance because, in a Penfield-type arm movement, the experience of acting is missing. This is different when one raises one’s arm automatically, say so as to catch a ball. The same applies to the examples on my list of automatic actions. When we shift gears, turn a doorknob, or unscrew the cap from the toothpaste, it certainly does not feel as if an external power is controlling our physical movements. Rather, in each of these cases, there is an experience of acting.

I do find Searle’s explanation of what an intention in action is helpful. However, if the intentionality of automatic actions is indeed to be equated with the experience of acting, or, in more general terms, the experience of agency, then it will remain mysterious why beliefs should fail to exhibit an analogous kind of intentionality. To argue that beliefs do not instantiate the experience of agency is to equate belief formation with the arm movement in Penfield cases, or with Tom’s twitching eye lid, or with the behavior of Peter’s paralyzed arm. But the claim that the phenomenology of belief is like the phenomenology of these kinds of involuntary behavior is exceedingly implausible. The involuntary behaviors just mentioned lack what typically characterizes our beliefs: reason responsiveness. When we form and drop beliefs, we respond to our evidence and are aware of being responsive to our evidence. Since we are aware of our ability to believe differently as our evidence changes, there is associated with belief an experience of agency that is also characteristic of automatic actions.

To sum up: Given that compatibilists conceive of voluntary control by distinguishing between good and bad causes, it follows that doxastic attitudes are under our voluntary control whenever the process by which they are formed exhibits reason responsiveness. In response, involuntarists can argue that behavior is under our voluntary control only to the extent it can be described as carrying out an intention. This move runs afoul of the fact that automatic actions are voluntary but cannot be so described. Attempting to handle the problem automatic actions pose, involuntarists might argue that such actions exhibit a kind of intentionality that doxastic attitudes lack. However, if that kind of intentionality is to be identified with Searle’s intention in action, i.e., with an experience of acting or agency, then this maneuver fails, for beliefs exhibit such intentionality no less than automatic actions do.
**Intentionally Held Belief**

As an alternative to claiming that there is a kind of intentionality that automatic actions have and beliefs do not, involuntarist could argue that voluntarily performed actions need not be intentional, but they need to be such that, in principle, they can be performed intentionally. Each of the examples I mentioned above—gear shifting, doorknob turning, the unscrewing of the toothpaste cap—can be done by way of carrying out an antecedently formed intention. It’s just that typically we do these things automatically, without having to dwell on them in advance. Doxastic attitudes, in contrast, cannot be adopted intentionally at all. In response to this modified argument from intentionality, the appeal to automatic actions in not effective. However, the modified argument fails if doxastic attitudes can be adopted intentionally. Next, I will argue that it is indeed possible to acquire a belief intentionally.

Here is a case in which I believe something because I intended to believe it. Suppose that, having returned from a trip and taken a shuttle to the airport parking garage, I am now where I thought I left my car. To my surprise, it is no longer there. I wonder whether it has been stolen. There is of course the possibility that I don’t accurately remember where I parked it. So I retrieve the paper slip which states the exact location of the parking spot I had chosen at the outset of my trip. According to the slip, I am at the right spot. Considering my evidence—the absence of my car and the parking slip—I conclude that my car was stolen. Since I do not suffer from a mental illness or a brain lesion preventing me from acquiring this belief, I encounter no problem executing my decision. Upon deciding to believe it, I instantly believe it. This, I submit, is a case in which I decided to believe that my car was stolen, and in which my belief is intentionally held. In believing that my car was stolen, I carry out the intention to believe this. Cases like this show that the second premise of the Argument from Intentionality is false as well.

This conclusion might be resisted because, in typical cases, a decision is momentary but the resulting intention persists. Suppose, after working in the garden, I decide to wash my hands because I want to prepare a meal. While I am on my way to the sink, a friend calls me on the phone. After the phone call is over, I finally wash my hands. My decision has been executed, and my intention has been carried out. Involuntarists might take this example to show that, by way of ϕ-ing, one
carries out an intention to \( \varphi \) only if the following condition is met: after the decision to \( \varphi \) has been made, there is a temporal interval, beginning with the time of the decision and ending with the time of the decision’s execution, during which there is an intention to \( \varphi \) that has not yet been carried out. The process leading to my belief that my car has stolen does not meet this condition, for the decision to believe this and the acquisition of the belief are simultaneous or are at least perceived as simultaneous. Hence in believing that my car was stolen, I do not carry out any intention.

However, a temporal interval as described above is not a necessary feature of forming an intention and then carrying it out. Suppose I decide to blink. If there is a temporal gap between my decision and the movement of my eye lid, it is too small for me to perceive it. In my experience, it seems that the decision and its execution are simultaneous. Nevertheless, it is correct to say that I intended to blink, and that by moving my eyelid down and up I carried this intention out. It’s not the case, therefore, that \( \varphi \)-ing intentionally always involves an interval during which one’s intention has not yet been carried out. Hence the demand for such an interval is not a good objection to my claim that I intentionally acquire the belief that my car was stolen.

The features of the car theft example can be summed up as follows:

(i) Wondering whether \( p \) is true, I suspend judgment about \( p \).
(ii) I consider my reasons for and against \( p \).
(iii) Concluding that I have good reasons for taking \( p \) to be true, I decide to believe \( p \).
(iv) My attitude of suspending judgment about \( p \) is replaced by that of believing \( p \).
(v) I believe \( p \) because I conclude that I have reasons for taking \( p \) to be true, and the causal relation between my decision to believe \( p \) and S’s believing \( p \) is non-deviant.26

25 One might hesitate to refer to blinking as an action. It might be better classified by calling it an instance of ‘intentional bodily behavior’. Note, however, that if by blinking I give someone a signal, we would be less hesitant to call it an action.

26 The non-deviancy clause is needed because there are possible cases in which S’s decision to believe \( p \) causes S’s belief that \( p \), but not in the right way. Deviant causation is well known when it comes to the relation between the intention to \( \varphi \) and an agent’s actual \( \varphi \)-ing. Suppose Ben intends to kill his uncle. While driving, he is thinking about how to get the job done. Being unnerved by his thoughts, Ben loses control over his car and runs over a pedestrian, killing him. The pedestrian happens to be Ben’s uncle. So Ben intended to kill his
These conditions, I submit, are sufficient for intentionally acquiring a belief. Suppose I decide to do something in an analogous fashion:

(a) I wonders whether I should ϕ.
(b) I consider my reasons for and against ϕ-ing.
(c) Concluding that I have good reason to ϕ, I decide to ϕ.
(d) I no longer wonder whether to ϕ but commence ϕ-ing.
(e) I am ϕ-ing because I decided to ϕ, and the causal relation between my decision to ϕ and my ϕ-ing is non-deviant.27

The degree of similarity between the belief case and the action case is striking. It is difficult to see why, if (a)-(e) are sufficient for ϕ-ing intentionally, (i)-(v) should not be sufficient for intentionally believing p.

One might wonder why so many authors are convinced that it is impossible to hold a belief intentionally. The answer, it seems to me, is that these authors think only of intentions as responses to practical reasons, overlooking the possibility that an intention to believe might be a response, not to practical reasons, but to epistemic reasons. It is this attitude that motivates the next argument I will discuss.

The Monetary Incentive Argument

Advocates of involuntarism have one option left to block the compatibilist case for doxastic voluntarism. They could concede that my belief that my car was stolen does exhibit a kind of intentionality—call it epistemic intentionality—but that that kind of intentionality isn’t the right kind needed for voluntary control. What voluntary control requires is practical intentionality, which is the kind of intentionality resulting from decisions that respond to practical reasons. The test whether ϕ-ing is something over which one has voluntary control is to ask oneself whether one can ϕ in response to a monetary reward. Thus the question at hand is: Is it possible to acquire a doxastic attitude in response to a monetary reward? I would agree that the answer to this question is ‘no’. Belief formation responds to our evidence (or

27 See the previous note.
sometimes what we mistakenly take to be our evidence), but never responds to
monetary rewards (perhaps neglecting rare cases).

The Monetary Incentive Argument can be stated as follows:

(1) If one has voluntary control over ϕ-ing, then one can ϕ in response
to a monetary reward.

(2) One cannot adopt a doxastic attitude in response to a monetary
reward.

Therefore:

(3) One does not have voluntary control over one’s doxastic attitudes.\(^{28}\)

The first premise of this argument is false. I conclude this paper by stating five
reasons for rejecting it:

(i) The Moral Fiber Argument: Jones is an American GI fighting the Germans in
the last stages of WW II. In battle, he deliberately and intentionally kills an enemy
soldier whose name is Schmidt. A man of strong moral fiber, it would have been
impossible for Jones to kill Schmidt in response to a monetary incentive. The first
premise implies that Jones’s killing Schmidt was not under Jones’s voluntary control.
To avoid this implausible outcome, the first premise must be rejected.

(ii) The St. Francis Argument. Actions can be performed in response to prudential
and in response to moral reasons. Suppose St. Francis performs a good deed in
response to a moral reason. He could not have performed this deed in response to
a monetary incentive. The 1st premise has the implausible consequence that St.
Francis’s good deed was involuntary.

(iii) The Argument from Arbitrariness: Actions can be performed in responsiveness
to moral or prudential reasons. According to the monetary incentive argument, ϕ-ing
in response to a moral reason is voluntary only if the agent could also ϕ in
response to a prudential reason. But assigning priority to prudential over moral
reasons is arbitrary. To avoid such arbitrariness, involuntarists could enlarge the
scope of the first premise by replacing the appeal to monetary incentives with an
appeal to practical reasons, understood broadly enough to include both prudential
and moral reasons. But then we may wonder why, when it comes to beliefs,
assigning priority to practical over epistemic reasons isn’t just as arbitrary as
assigning priority to prudential over moral reasons. If we aim at a general account

\(^{28}\) For deployment of this argument, see Bennett 1990 and Chrisman 2008.
of voluntary control that avoids arbitrary favoritism of one kind of reason over another, then we must reject the first premise.

(iv) *The Kinds of Control Argument*: Since one can be responsive to one kind of reason without, at the same time, being responsive to another kind of reason, it makes sense to distinguish between as many kinds of voluntary control as there are kinds of reasons to which we might or might not be responsive. Let us distinguish between three: Prudential control, moral control, and epistemic control. Suppose St. Francis ϕ-s, and, while his ϕ-ing is an indication of responsiveness to moral reasons, it does not involve any responsiveness to prudential reasons. The first premise yields the simplistic and implausible result that his ϕ-ing was not under his control at all. A better approach is to say that, when St. Francis ϕ-s, he has moral but not prudential control over ϕ-ing. Since St. Francis has moral control over his ϕ-ing, it would be bizarre to deny that he bears responsibility for ϕ-ing and can be given credit for it. When it comes to control over beliefs, an analogous treatment is called for. In believing that you have hands, you are responsive to epistemic reasons but not to practical reasons. Though you do not have practical control over the belief, you enjoy epistemic control. It would be equally bizarre to deny that you bear responsibility for it, and that you can be given credit for believing in accord with your evidence.

(v) *The Missing Argument Argument*: According to the first premise of the Monetary Incentive Argument, responsiveness to practical reasons is a necessary condition of voluntary control over both actions and beliefs. In light of (i)–(iv), we may conclude that this premise is not self-evident. Since it is not self-evident, unless the premise is defended with a good argument, it is unacceptably *ad hoc*. Alas, thus far I have not seen any argument in its support at all.29

---

29 For helpful comments or discussion, I wish to thank Mylan Engel, Richard Fumerton, Adam Leite, Bruce Russell, and Stefan Tolskdford.
Bibliography


