

## Deontology and Descartes' Demon

Brian Weatherson, Cornell University

January 2007

### 1. *Digesting Evidence*

In his *Principles of Philosophy*, Descartes says,

Finally, it is so evident that we are possessed of a free will that can give or withhold its assent, that this may be counted as one of the first and most ordinary notions that are found innately in us. (Descartes 1644/1965: xxix)

In this paper, I'm going to defend a broadly Cartesian position about doxastic freedom. At least some of our beliefs are freely formed, so we are responsible for them. Moreover, this has consequences for epistemology. But the *some* here is crucial. Some of our beliefs are not freely formed, and we are not responsible for those. And that has epistemological consequences too. Out of these considerations a concept of doxastic responsibility arises that is useful to the *externalist* in responding to several challenges. I will say at some length how it helps with the New Evil Demon problem, and I'll note some difficulties in reconciling internalism with the idea that justification is a kind of blamelessness. The internalist, I'll argue, has to say that justification is a kind of praiseworthiness, and this idea that praise is more relevant to epistemic concepts than blame will be a recurring theme of the paper.

While the kind of position I'm adopting has been gaining supporters in recent years, it is still largely unpopular. The arguments of William Alston (1988) have convinced many that it is a mistake to talk of doxastic freedom, or doxastic responsibility. The short version of this argument is that our beliefs are involuntary, and freedom and responsibility require voluntariness. The longer, and more careful, argument involves drawing some distinctions between ways in which we might come to be in a state. It helps to start with an example where the normative facts are relatively uncontroversial, namely digestion.

Imagine that Emma eats a meat pie, and due to a malfunction in her stomach the pie is not properly digested, leading to some medical complications. Is Emma responsible for her ill-health? Well, that depends on the back-story. If Emma knew that she could not properly digest meat pies, but ate one anyway, she is responsible for the illness via her responsibility for eating the pie. Even if Emma did not know this, she might be responsible for the state of her stomach. If her stomach could not digest the pie because it had been damaged by Emma's dietary habits, and let's say Emma knew that her diet could damage her stomach, then Emma is responsible for the state of her stomach and hence for the misdigestion of the pie and hence for her ill-health. But if neither of these conditions obtain, if it just happens that her stomach misdigests the pie, then Emma is not responsible for her ill-health. Even though the cause of her ill-health is something that *her* stomach does, she is not responsible for that since her stomach is not under her voluntary

control. Put another way, her responsibility for maintaining her own health means that she is responsible for the type of digester she is, but he is not responsible for this token digestion.

Simplifying a little, Alston thinks that the case of belief is similar. Say that Emma has a false belief that  $p$ . Is she responsible for this piece of doxastic ill-health? Again, that depends on the back story. If Emma believes that  $p$  because she was careless in gathering evidence, and the evidence would have pointed to  $\sim p$ , then she is responsible for being a bad gatherer of evidence. If Emma has been negligent in maintaining her doxastic health, or worse if she has been doing things she knows endangers doxastic health, then she is responsible for being the type of believer she is. But she is never responsible merely for the token belief that is formed. Her mind simply digests the evidence she has, and Emma's responsibility only extends to her duty to gather evidence for it, and her duty to keep her mind in good working order. She is not responsible for particular acts of evidential digestion.

But these particular acts of evidential digestion are the primary subject matters of epistemology. When we say Emma's belief is justified or unjustified, we usually mean that it is a good or bad response to the evidence in the circumstances. (I'm obviously here glossing over enormous disputes about what makes for a good response, what is evidence, and what relevance the circumstances have. But most theories of justification can be fit into this broad schema, provided we are liberal enough in interpreting the terms 'good', 'evidence' and 'circumstances'.) If Emma is not responsible for her response to the evidence, then either we have to divorce justification from responsibility, or we have to say that the concept of justification is defective.

We can summarise these considerations as a short argument. The following formulation is from Sharon Ryan (2003: 48).

1. If we have any epistemic obligations, then doxastic attitudes must sometimes be under our voluntary control.
2. Doxastic attitudes are never under our voluntarily control.

---

3. We do not have any epistemic obligations.

Ryan goes on to reject both premises. Matthias Steup (2000, 2005) also rejects both premises of this argument. I'm more sympathetic to premise 1, but I (tentatively) agree with them, against what sometimes seems to be orthodoxy, that premise 2 fails. That is, I endorse a kind of doxastic voluntarism. (Just what kind will become clearer as we go along.) There are four questions that anyone who endorses voluntarism, and wants to argue that this matters epistemologically, should I think answer. These are:

- (A) What is wrong with current arguments against voluntarism?
- (B) What does the voluntariness of (some) beliefs consist in?
- (C) Which kinds of beliefs are voluntary?
- (D) What difference does the distinction between these classes make for epistemology?

My answer to (A) will be similar to Ryan's, and to Steup's, but with I think enough differences in emphasis to be worth working through. My answer to (B), however, will be a little more different. I'm going to draw on some work on self-control to argue that some beliefs are voluntary because they are the result of exercises of, or failures to exercise, self-control. My answer to (C) is that what I'll call *inferential* beliefs are voluntary, while *perceptual* beliefs are not. Ryan and Steup sometimes seem to suggest that even perceptual beliefs are voluntary, and I don't think this is true. The consequence for this, I'll argue in answering (D), is that inferential beliefs should be judged by how well they respond to the evidence, while perceptual beliefs should be judged by how well they reflect reality.

I'll offer my answers to those four questions in sections 2, 4, 6 and 7. In between I'll discuss recent work on self-control (section 3) and the contrast between my answer to (B) and other voluntarist answers (section 5). In section 8 I'll say how my partially voluntarist position gives the externalist a way to avoid the New Evil Demon problem. And in section 9 I'll make a direct argument for the idea that justification is a kind of praiseworthiness, not a kind of blamelessness.

Before we start, I want to note two ways, other than Ryan's, of formulating an argument against doxastic responsibility. These are going to seem quite similar to Ryan's formulation, but I think they hide important differences. The first version uses the idea that some doings (or states) are volitional. That is, we do them (or are in them) because we formed a volition to do so, and this volition causes the doing (or state) in the right kind of way.

1. If we have any epistemic obligations, then either the formation or maintenance of doxastic attitudes must sometimes be volitional.
  2. The formation or maintenance of doxastic attitudes is never volitional.
- 
3. We do not have any epistemic obligations.

I won't argue against premise 2 of this argument, though Carl Ginet (1985, 2001) has done so. But I think there's little to be said for premise 1. The principle behind it is that we are only responsible for volitional doings. And that principle is very dubious. We could run the kind of regress arguments against it that Gilbert Ryle (1949) offers. But it is simpler to note some everyday counterexamples. Borrowing an example from Angela M Smith (2005), if I forget a friend's birthday, that is something I am responsible and blameworthy for, but forgetting a

birthday is not volitional. (Below I'll offer a Rylean argument that we are sometimes praiseworthy for doings that are not volitional.) So this argument fails. Alternatively, we could run the argument by appeal to freedom.

1. If we have any epistemic obligations, then doxastic attitudes must sometimes be free.
  2. Doxastic attitudes are never free.
- 
3. We do not have any epistemic obligations.

Premise 1 of this argument is more plausible. But, as we'll see presently, premise 2 is not very plausible. Whether Descartes was right that premise 2 is obviously false, it does seem on reflection very hard to defend. So this argument fails. Ryan's formulation is interesting because it isn't clear just which of the premises fails. As I said, I'm going to suggest that premise 2 fails, and that doxastic attitudes are voluntary. But this will turn on some fine judgments about the voluntary/involuntary boundary. If I'm wrong about those judgments, then the arguments below will suggest that premise 1, not premise 2, in Ryan's formulation fails. Either way though, the argument is unsuccessful.

## *2. Responding to the Involuntarists*

There are two kinds of argument against the idea that belief is voluntary. One kind, tracing back to Bernard Williams (1973), holds that the possibility of voluntary belief can be shown to be incoherent by reflection on the concept of belief. This argument is no longer widely endorsed. Nishi Shah (2002) provides an excellent discussion of the problems with Williams' argument. I will focus on the other kind, that claims we can see that belief is involuntary by observing differences between beliefs and paradigm cases of voluntary actions. I will make three objections to these arguments. First, the argument looks much less plausible once we distinguish between having a belief and forming a belief. Second, the argument seems to rely on inferring from the fact that we don't do something (in particular, believe something that we have excellent evidence is false) to the conclusion that we can't do it. As Sharon Ryan (2003) points out, it might simply be that we won't do it. Third, the argument relies on too narrow a conception of what is voluntary, and when we get a more accurate grasp on that concept, we'll give up the argument. Here is a representative version of the argument from William Alston.

Can you, at this moment, start to believe that the United States is still a colony of Great Britain, just by deciding to do so? ... [S]uppose that someone offers you \$500,000,000 to believe it, and you are much more interested in the money than in believing the truth. Could you do what it takes to get that reward? ... Can you switch propositional attitudes toward that proposition just by deciding to do so? It

seems clear to me that I have no such power. Volitions, decisions, or choosings don't hook up with anything in the way of propositional attitude inauguration, just as they don't hook up with the secretion of gastric juices or cell metabolism. (Alston 1989: 263)

Now Alston does note, just one page earlier, that what is really relevant is whether our being in a state of belief is voluntary, not whether the activity of belief formation is voluntary. But he thinks nevertheless that issues about whether we can form beliefs, any old beliefs it seems, voluntarily matters to the question about the voluntariness of belief states.

If we think about what it is to be in a state voluntarily, this all seems beside the point. We can see this by considering what it is to be in a political state voluntarily. Consider Shane, who was born into Victoria. His coming to be in Victoria was hence not, in any way, voluntary. Shane is now a grown man, and he has heard many travellers' tales of far away lands. But the apparent attractions of Sydney and other places have no pull on Shane; he has decided to stay in Victoria. If he has the capacity to leave Victoria, then Shane's continued presence in Victoria is voluntary. Similarly, we are voluntarily in a belief state if we have the capacity to leave it, but choose not to exercise this capacity. Whether the belief was formed voluntarily is beside the point.

If Shane leaves a state, the natural place to leave is for another state, perhaps New South Wales or South Australia. It might be thought that if we leave a belief state, we have to move into another belief state. So to have this capacity to leave, we need the ability to form beliefs voluntarily. Not at all. The capacity to become uncertain is capacity enough. (If Shane has a boat, and the capacity to flourish at sea, then perhaps he too can have the capacity to leave Victoria without the capacity to go into another state.)

But do we have the capacity to become uncertain? Descartes thought so; the point of the First Meditation is to show us how to exercise this capacity. Moreover, this capacity need not be one that we exercise in any particularly nearby possible worlds. We might exercise our freedom by always doing the right thing. As Descartes goes on to say in the Fourth Meditation.

For in order to be free, there is no need for me to be capable of going in each of two directions; on the contrary, the more I incline in one direction – either because I clearly understand that reasons of truth and goodness point that way, or because of a divinely produced disposition of my inmost thoughts – the freer is my choice. (Descartes 1641/1996: 40)

This seems like an important truth. A roulette wheel is not free, and nor is where it ends up a matter of its own voluntary control. If we want to be free, we should not aim to resemble a roulette wheel. John Cottingham (2002) interprets Descartes here as suggesting that there is a gap between free action and voluntary action, contrasting his “strongly compatibilist notion of human

freedom” (350) with the “doxastic involuntarism” (355) suggested by the following lines of the Third Meditation.

Yet when I turn to the things themselves which I think I perceive very clearly, I am so convinced by them that I spontaneously declare: let whoever can do so deceive me, he will never bring it about that I am nothing, so long as I continue to think that I am something ... (Descartes 1641/1996: 25)

Now there are two questions here. The first is whether Descartes intended to draw this distinction. That is, whether Descartes thought that the kind of free actions that he discusses in the Fourth Meditations, the free action where we are incapable of going in the other directions, are nevertheless involuntary. I don't have any informed opinions about this question. The second is whether this kind of consideration supports the distinction between the free and the voluntary. And it seems to me that it does not. Just as Descartes says the free person will be moved by reasons in the right way, it seems natural to say that a person who acts voluntarily will be responsive to reasons. Voluntary action does require freedom from certain kinds of coercion, but the world does not coerce us when it gives us reason to believe one thing rather than another. If we have voluntary control over our beliefs, then we should be compelled by the sight of rain to believe it is raining.

In her discussion of the puzzle of imaginative resistance, Tamar Szabó Gendler (2000) notes that philosophers have a tendency to read too much into intuitions about certain cases. What we can tell from various thought experiments is that in certain circumstances we won't do a certain thing. But getting from what we won't do to what we can't do is a tricky matter, and it is a bad mistake to infer from won't to can't too quickly. Matthias Steup (2000) points out that if we try to stick a knife into our hand, we similarly won't do it. But this is no evidence that we cannot do it. And Sharon Ryan (2003) notes that we will not bring ourselves to run over pedestrians for no reason. For most of us, our moral sense prevents acting quite this destructively. Yet our continued avoiding of pedestrians is a series of free, even voluntary, actions. We could run over the pedestrians, but we will not. Since forming false beliefs is a form of self-harm, it is not surprising that it has a similar phenomenology, even if it is genuinely possible.

It might be argued that we will engage in small forms of self-harm that we can do when the financial rewards are great enough. So we should be able to form this belief about the United States for a large amount sum of money. But I suspect that the only way to exercise the capacity to believe the United States is still a colony is by first suspending my belief that it is no longer a colony. And the only way I can do that is by generally becoming more sceptical of what I've been told over the years. Once I get into such a sceptical mood, I will be sceptical of claims that I'll get half a billion dollars should I have this wild political belief. So I won't form the belief in part because the 'promisor' lacks the capacity to sufficiently convince me that I'll be richly rewarded for doing so. This looks like a lack of capacity on their part, not my part.

The final point to make about this argument, and those like it, is that if we are to conclude that belief formation is *never* voluntary, then we need to compare it to all kinds of voluntary action. And Alston really only ever compares belief formation to volitional action. If this doesn't exhaust the range of voluntary action, then belief formation might be properly analogous to some other voluntary action. Indeed, this turns out to be the case. To see so, we need to make a small detour through modern work on self-control.

### *3. How to Control Your Temper*

To start, let's consider three examples of a person failing to keep a commitment they've made about what the good life is. The three ways will be familiar from Gary Watson's discussion of recklessness, weakness and compulsion (Watson 1977), and the discussion of these cases by Jeanette Kennett and Michael Smith (1994, 1996). My characterisation of the cases will turn out to differ a little from theirs, but the cases are similar. Each of the examples concerns a character Murray, who has decided that he should not swear around his young son Red. He resolves to do this, and has been working on curbing his tendency to swear whenever anything bad happens. But three times over the course of the day he breaks his commitment.<sup>1</sup>

The first time comes when Murray puts his hand down on a hot plate that he did not realise was on. The searing pain undermines his self-control, and he is unable to stop himself from swearing loudly through the pain.

The second time comes when Murray drops and breaks a wine glass. Murray doesn't lose his self-control, but he doesn't exercise the self-control he has. He temporarily forgets his commitment and so, quite literally, curses his misfortune. On doing so he immediately remembers that Red is around, and the commitment he has made, and regrets what he did.

The third time comes on the tram home, when Murray gets into a disagreement with a political opponent. Murray can't find the words to express what he feels about the opponent without breaking his commitment. So he decides, without much reason, that his need to express what he feels outweighs his commitment, and starts describing his opponent using language he would, all things considered, not have used around young Red.

The first and third cases are close to textbook cases of compulsion and recklessness. Note in the first case that when Murray reflects back on what happened, he might be irritated that his work on reducing his tendency to swear has not been more successful. But he will not be upset that he didn't exercise more self-control on that occasion. He didn't have, no normal person would have, the amount of self-control he would have needed to stop swearing then. All that would help is having the disposition to say different things when his self-control is defeated. And that's not a disposition he can acquire on the spot.

---

<sup>1</sup> The cases, especially the second, were inspired by Richard Holton's (2003, 2004) discussion of resolutions to prevent 'automatic' actions like smoking or sleeping in.

I've described the first case as one where Murray's self-control is undermined. This is a term taken from recent work by Richard Holton and Stephen Shute (forthcoming), who carefully distinguish between self-control being *undermined* by a provocation, and it being *overwhelmed* by a provocation. Undermining occurs when the provocation causes the agent to have less self-control than they usually have; overwhelming occurs when the provocation is too much for the agent's control. The difference is relevant to them, because they are interested in what it is for an agent to *lose* control. That seems to be what happens here. After all, the things one would naturally do afterwards (jumping around, screaming, swearing if one's so disposed) don't seem particularly controlled by any measure.

Similarly I've accepted Watson's description of cases like the third as instances of recklessness, but we shouldn't think this necessarily contrasts with weakness. It might be that in this case Murray is both weak and reckless. He is not akratic, if we stipulatively define akrasia as acting against one's better judgment. But if we accept Richard Holton's view that weakness of will consists in being "too ready to reconsider their intentions" (page 241), then in this case Murray is weak-willed.<sup>2</sup> This seems to be the right way to talk about the case to me. With these details in place, we can talk about what's crucial to this essay, the contrast with the second case.

In the second case Murray fails to exercise self-control. He could have prevented himself from swearing in front of his son. Breaking a wine glass is irritating, but it neither undermines nor, necessarily, overwhelms self-control. Murray had the capacity to think about his resolution to not swear in front of Red. And if he had exercised this capacity, he would not have sworn when he did.

In the first case, Murray will only regret his lack of prior work at changing his dispositions in cases where his control fails. In the second case he will regret that, but he will also regret what he did on that occasion, for he could have kept his resolution, had only he thought of it. This regret seems appropriate, for in the second case he did something wrong at the time he swore, as well perhaps as having done something wrong earlier. (Namely, not having worked hard enough on his dispositions.) This difference in regret doesn't constitute the difference between compulsion and a case where self-control fails, but it is pretty good evidence that this is a failure of self-control.

So the second case is not one where Murray was compelled. He had the capacity to keep his commitment, and nothing was stopping him exercising this control, but he failed to do so. His

---

<sup>2</sup> Whether Murray is akratic is a slightly more complicated question than I've suggested in the text. If akrasia is acting against one's judgment, then he is not; if akrasia is acting against one's *considered* judgment, then he is. 'Akrasia' is a technical term, so I don't think a huge amount turns on what we say about this question.

There is an interesting historical precedent for Holton's theory of weakness of will. Ryle hints at a similar position to Holton's when he says "Strength of will is a propensity the exercise of which consist in sticking to tasks' that is, in not being deterred or diverted. Weakness of will is having too little of this propensity." (1949: 73) But the idea is not well developed in Ryle. We'll return below to the differences between Ryle's and Holton's theories.

failure was a failure of self-control. Murray's self-control is, in this case, overwhelmed by the provocation. But it need not have been. Within some fairly broad limits, how much self-control we exercise is up to us.<sup>3</sup> Murray's failure of self-control is culpable because anyone with the capacity for self-control Murray has could have avoided breaking his commitment. I'm not going to try to offer an analysis of what it is to have a capacity, but I suspect something like the complicated counterfactual analysis Kennett and Smith offer, and that Smith offers elsewhere (Smith 1997, 2003), is broadly correct.<sup>4</sup>

Kennett and Smith stress two things about this capacity that are worth noting here. First, having this kind of capacity is part of what it is to be rational. That is, being rational requires thinking of the right thing at the right time. As Ryle says, "Intelligently reflecting how to act is, among other things, considering what is pertinent and disregarding what is inappropriate." (1949: 31) Second, Kennett and Smith note that exercises of this capacity cannot be volitional. Following Davidson (1963), they say they cannot be *actions*. I find this terminology somewhat strained. Catching a fast moving ball is an action, I'd say, but it doesn't seem to be volitional. So I'll use 'volitional action' for this Davidsonian sense of action.

Many recent philosophers have endorsed the idea that some of the mental states for which we hold people responsible are not voluntary, or at least are not volitional. Adams (1985), Heller (2000), Owens (2000: Ch. 8) and Hieronymi (forthcoming) note ways in which we appropriately blame people for being in certain states, where being in that state is not volitional. Something like this idea seems to be behind Ryle's several regress arguments against the intellectualist legend. It just isn't true that what we do divides cleanly into outcomes of conscious thought on the one hand, and mere bodily movements (a la digestion) on the other.<sup>5</sup> Rather there is a spectrum of cases from pure ratiocination at one end to pure bodily movement at the other. And some of the things in the middle of this spectrum are proper subjects of reactive attitudes. The focus in this literature has been on blame, but some states in the middle of this spectrum are also praiseworthy.

Consider some action that is strikingly imaginative, e.g. a writer's apt metaphor or, say, a cricket captain's imaginative field placements. It seems that, assuming the field settings are successful, the captain deserves praise for being so imaginative. But of course the captain did not, really could not, first intend to imagine such field settings, then carry out that intention. So something for which the captain deserves praise, his act of imagination, is not volitional. So not all praiseworthy things we do are volitional.

There are two responses to this argument that I can imagine, neither of them particularly plausible. First, we might think that the captain's imagination is simply a remarkable feature of

---

<sup>3</sup> Holton (2003) compares self-control to a muscle that we can exercise. We can make a similar point to the one in the text about physical muscles. If I try to lift a box of books and fail, that doesn't show I lack the muscular capacity to lift the box; I might not have been trying hard enough.

<sup>4</sup> Ryle (1949: 71ff) also offers a counterfactual account of capacities that seems largely accurate.

<sup>5</sup> As I read him, Ryle takes this fact to reveal an important weakness in Descartes' theory of mind.

































