

Free Will Agnosticismⁱ

1. Introduction

In recent years, many interesting theses about free will have been proposed that go beyond the compatibilism/incompatibilism debate. Semicompatibilists believe that (the freedom relevant to) moral responsibility is compatible with determinism even if the freedom to do otherwise is not (see Fischer 2006, Fischer and Ravizza 1998). Revisionist compatibilists believe that free will is compatible with determinism, but our concept of free will needs significant revision (Vargas 2009). Hard incompatibilists claim that free will is incompatible with determinism and with any world (including ours) that does not involve agent-causation (Pereboom 2001). Impossibilists, who believe that free will and moral responsibility are impossible (Strawson 1991). Free will contextualists believe that sentences expressing attributions of freedom to agents may differ in truth value depending on the context of the utterer of the sentence (Rieber 2006). Agnostic autonomists stay neutral about whether or not free will is compatible with determinism, but claim that it is more likely that we have free will than that we do not (Mele 1995). Non-cognitivists claim that sentences attributing freedom to agents are not truth-apt, but rather express certain attitudes (Honderich 2002). Free will illusionists claim that there are important senses in which we lack free will (namely, we lack ‘libertarian free will’), but it is important that we remain under the illusion that we have it (Smilansky 2002). In this paper, I wish to formulate and defend an overlooked, but significant and plausible non-standard view of free will, which I shall call ‘free will agnosticism’ (FWA).ⁱⁱ

2. Weak FWA

A weak form of agnosticism about free will states:

Weak FWA: No one knows whether or not normally functioning adult humans have free will.

By ‘no one’ I mean no human who presently exists, or has existed. After arguing for this version of FWA, I shall examine stronger versions to see if they are also plausible. Before I begin my argument for Weak FWA, it is worth making two points about what it claims.

First, perhaps any formulation of FWA will remain ambiguous unless it is specified what ‘free will’ means. That is, perhaps ‘free will’ is ambiguous. Alternatively, there may be differing *conceptions* of what free will is, and no version of FWA is fully explicit unless it tells us what conception of free will it is assuming. Lastly, it might be that there are different *types* of free will. Whether one takes the following to be different concepts, types or conceptions of free will, there are three main ways to understand free will. Free will is often considered the ability to choose options from various possible alternatives (let us call this ‘freedom to do otherwise’). It is also seen to be a capacity to originate one’s actions (‘sourcehood freedom’). Lastly, it is often thought of as the set of abilities and capacities one needs in order to be a morally responsible agent (‘MR-freedom’). Whatever the relationship between these notions (perhaps one is the definition of free will, while another conveys its nature), my argument for FWA applies to all of them (see section 3.1.2. below).

Second, contextualists about knowledge will claim that any sentence purporting to express FWA will express different propositions in different contexts due to the context-sensitivity of the word ‘know’ (see, for example, Lewis 1996). Thus one standard position would be that, in the epistemology class room, where consideration of skeptical scenarios is prominent, an utterance

of “No one knows whether there is free will” expresses a truth, while in less demanding everyday contexts, an utterance of the same sentence expresses a falsehood. I do not have space to discuss contextualism and the impact it has on FWA. I shall simply assume that ‘know’ is not context sensitive. If it is, I intend to be using it in an everyday, low standard way.ⁱⁱⁱ

3. An Argument for Weak FWA

In this section, I will present an argument that we do not know whether there is free will (I shall sometimes use ‘we do not know’ to mean ‘humans past and present do not know’, and use ‘we have free will’ or ‘there is free will’ to mean ‘normally functioning human adults have free will’). My argument for this thesis is divided into two parts. The first part is an argument (in fact, two arguments) for the conclusion that we do not know that we have free will. The second part is an argument for the conclusion that we do not know that we lack free will. The conjunction of these conclusions is equivalent to the claim we do not know whether or not we have free will.

3.1. The first argument that we don’t know we have free will

In this subsection and the next, I will argue that no one knows that normal adult humans have free will. The basic idea of both arguments is that the philosophical and scientific challenges to the existence of free will, though perhaps not strong enough to show that there is no free will, are strong enough to prevent us knowing that there is free will.

The challenge to free will we shall concentrate on is (a variation of) the classic problem of free will. Incompatibilists claim that determinism rules out free will, at least in worlds (like ours) that have a remote past, i.e. a past that precedes the existence of any human (this caveat is added to deal with the point in Campbell 2007 that certain arguments for incompatibilism fail because some deterministic worlds have no remote past). If these incompatibilists are right, not any old

indeterministic world with a remote past will be hospitable to free will (if, indeed, any are). Only those in which certain of our actions, or their immediate causal antecedents, are undetermined will even be candidates for being hospitable to free will.

Let us call the thesis that all our actions and their immediate causal antecedents are determined “action determinism”.^{iv} Let us call the thesis that some of our actions or their immediate causal antecedents are undetermined “action indeterminism”. If determinism rules out free will in worlds like ours, then so does action determinism. Now consider the following claim:

(A) Action determinism is true and incompatibilism is true.

Understand “action determinism” as above. Further, let us understand “incompatibilism” to mean the weak thesis that action indeterminism is required for the existence of free will in worlds like ours (i.e. worlds that have a remote past), and let us understand “compatibilism” to be the negation of this thesis. (A) says, then, that action determinism is true, and our having free will in this world requires that it is not true. (A) thus entails that there is no free will. Conversely, if there is free will, then (A) is false. Given this, a simple argument for FWA runs as follows:

- (1) If anyone knows there is free will, then she knows (or could easily work out) that (A) is false.
- (2) No one knows (or could easily work out) that (A) is false.
- (3) Therefore, no one knows there is free will.

The argument is valid. If there is a problem, then, it is with one or both of the premises.

3.1.1. The argument for (1)

Premise (1) is relatively easy to defend. We know that *if* there is free will, then (A) is false. If we further *know* that there is free will, then we could easily deduce that (A) is false from this knowledge plus our knowledge of the conditional. We have either made such a deduction, or could easily do so (i.e. there are no epistemic impediments to our doing so). If we were to do so, then (given that we know we have free will) we would come to know that (A) is false. Therefore, (1) is true. If anyone knows there is free will, they can easily deduce from their total knowledge that (A) is false.

In the above defense of (1), I am appealing to the idea that we can extend our knowledge through competently deducing propositions from known facts. I do not wish here to commit to any particular kind of closure principle on knowledge. All I need to get my argument off the ground is that in these particular instances it is that, if we know we have free will, we can come to know the negation of (A). Even the staunchest skeptic about closure principles must concede that we are at least sometimes able to expand our knowledge by such simple deduction.^v

3.1.2. The argument for (2)

The main work in the argument is being done by (2). In effect, (2) says that, for all we know, our actions (and their immediate causal antecedents) are determined and our having free will requires that at least some of our actions (or their immediate causal antecedents) are not determined. Why think this is true? A partial defense of (2) is that we know the falsity of neither conjunct of (A). That is, we do not know that action determinism is false, and nor do we know that incompatibilism is false. Therefore, we cannot know that (A) is false by deducing its falsity from the falsity of one of its conjuncts. Let us examine further the claim that we know the falsity of neither conjunct.

Obviously, no incompatibilist will say that we know that incompatibilism is false. The same applies to those who refrain from judging whether free will is compatible with determinism. Even many compatibilists would not be willing to say that we *know* free will is compatible with action determinism (in a world with a remote past). Though they may believe it is, many would fall short of claiming that they know it is compatible.

There is good reason for this. The arguments for incompatibilism, while not knockdown, are strong. The arguments for incompatibilism differ somewhat depending on how we conceive of free will. If free will is thought of as the freedom to do otherwise, then clearly the most prominent argument for incompatibilism is the Consequence Argument (see Van Inwagen 1983). This argument has many variations, but the basic idea is simple. If determinism is true, then the remote past and the laws of nature determine all our actions. If we have the freedom to do otherwise in such a deterministic world, then, we are able to do things such that, were we to do them, the past or the laws would be different. We are not able to do things such that, were we to do them, the past or the laws would be different. Therefore, if determinism is true, we do not have the freedom to do otherwise. Furthermore, as I have said above, if determinism rules out free will so does action determinism. Not any indeterministic world should satisfy the incompatibilist.

The Consequence Argument has been widely criticized (Lewis 1981, Kapitan 2002) and widely defended (O'Connor 2002, Huemer 2000). No consensus has been reached about its soundness. Even most opponents agree, however, that it represents an intuitively compelling line of reasoning. Indeed, many compatibilists are inclined to accept its main conclusion. Such (semi)compatibilists downplay the importance of the freedom to do otherwise in our thinking about free will, claiming that, though this freedom is incompatible with determinism, sourcehood

freedom and MR-freedom are compatible with determinism. Even those philosophers who believe that the freedom to do otherwise and determinism are compatible are aware of the complexity and subtlety of the issues at stake and would be foolhardy to claim that it is known that the freedom to do otherwise is compatible with determinism.

If one conceives of free will as sourcehood freedom, the arguments for incompatibilism are slightly different. An indirect argument is that, in order to be the source of one's actions, one must have the (incompatibilist) ability to do otherwise. One simple direct argument states that, in order to be the source of one's actions (and thus have free will), it must not be the case that any deterministic causal chain that produces one's actions can be traced to a source outside of one's control. Thus any deterministic causal chain that leads to free actions must start at an appropriate location and time within the agent that performs these actions (i.e. the chain must start with one of her actions or with its immediate causal antecedents). If action determinism is true, all an agent's actions and their immediate causal antecedents are deterministically caused. Therefore, sourcehood freedom is incompatible with determinism. Obviously compatibilists will deny that this is part of what it is to be the source of an action, but it must be admitted it is a natural, intuitive and simple conception of origination.

Furthermore, other arguments for incompatibilism do not simply assume this idea (that to be the source of our actions, these actions must not be deterministically caused by factors beyond our control). For example, Pereboom's four-case manipulation argument starts with the premise that, in certain carefully described cases of an agent's being manipulated, the agent is not free (see, for example, his 2001, 2005, and 2008). These cases are then gradually altered to make them more like cases of simple determinism. The final case is an example of an agent who is not manipulated but lives in a deterministic world. Pereboom argues that we should judge the

manipulated agents to be unfree and that there is no relevant difference (as regards free will) between the manipulated agents and the determined agent. He also argues that the best explanation for the manipulated agents not being free is precisely that the manipulated agent's actions are causally determined by factors beyond their control. This being so, a determined agent's actions are also not free. Just as (certain) manipulated agents are not the source of their actions, neither are determined agents the source of theirs. Pereboom's argument is one of many manipulation arguments for incompatibilism (see Kane 1996, Mele 1995 for others). Though there are many replies to these arguments in the literature (see, for example, McKenna 2008, Kearns 2012) this style of argument may be sound.

Many of the arguments for incompatibilism concerning MR-freedom are based on the arguments above. It is a very attractive thesis that either the freedom to do otherwise or sourcehood freedom is necessary for moral responsibility. Thus if both are incompatible with determinism, so is moral responsibility (and MR-freedom). The arguments above can be applied to MR-freedom more directly. Indeed, in defending his four-case argument, Pereboom explicitly states that he understands free will as MR-freedom. Just as we judge Pereboom's agents to be unfree, we also judge them to not be morally responsible for their actions. Lastly, the Direct Argument, an argument identical in form to the Consequence Argument, concludes that moral responsibility is incompatible with determinism (see van Inwagen 1983, for some replies see Kearns 2011, McKenna 2008, Haji 2008). The idea is that we are not responsible for the distant past or the laws of nature. Furthermore, if we are not responsible for these things, we are not responsible for anything they entail. If determinism is true, they entail every fact, including those describing our actions. Thus if determinism is true, we are not responsible for anything. Taken together, these arguments are strong enough to show that we do not know that MR-freedom is compatible with

determinism. And if it is not compatible with determinism, it is not compatible with action determinism. In essence, then, no matter how we understand free will, no one knows that incompatibilism is false.

It is even clearer that we do not know that action determinism is false. First, we do not know if any undetermined events occur at all. Whether determinism is false or not is still an open scientific question. Second, even if it was clear that some undetermined events occur, it is beyond our current ken whether such events occur at appropriate moments and appropriate locations in our brains.

We therefore know neither that action determinism is false, nor that incompatibilism is false. We do not know their conjunction is false by knowing one of the conjuncts is false. Of course, this is entirely consistent with our knowing that the conjunction is false for some other reason. Consider this conjunction: Goldbach's conjecture is true and Goldbach's conjecture is false. We do not know of either conjunct that it is false, but we do know that the conjunction is false. It is, after all, a contradiction.

Still, (A) is not a contradiction. Its conjuncts are controversial scientific and philosophical claims respectively. Furthermore, their conjunction seems to retain their controversy. We do not know that (A) is false through our current scientific knowledge, nor through our current philosophical knowledge. We do not know that (A) is false *a priori*. We do not know it is false directly from experience (or otherwise non-inferentially). And that (A) is false does not seem to be a consequence of our stock of commonsense everyday knowledge. But these are the only ways we might know that (A) is false. Thus we do not know it is.

3.1.3. An objection to (2)

I take it that the most controversial aspect of the above argument for (2) is the claim that the falsity of (A) is not a consequence of our everyday commonsense knowledge. After all, that (A) is false is a consequence of the proposition that we have free will. Thus if we know that we have free will, we can easily work out that (A) is false (this is simply premise (1)). Furthermore (it might be claimed), that we have free will is part of our everyday commonsense knowledge.

This is, in effect, the modus ponens to my modus tollens. I claim that if we know we have free will, we know (A) is false (premise (1)). We don't know (A) is false (premise (2)), therefore we don't know we have free will. The claim under consideration agrees with the conditional (i.e. premise (1)), asserts that we do know we have free will, and concludes that we know (A) is false.

This response is reminiscent of a Moorean response to skepticism (Moore 1939). We each know that we have hands, and thus we know (or are in a position to know) the proposition that we are not handless brains in vats. I am sympathetic with this response to skepticism. However, the fact that we have hands is a Moorean fact (a fact which “is one of those things that we know better than we know the premises of any philosophical argument to the contrary” (Lewis 1996)). It is part of our commonsense body of knowledge. The proposition that we have free will is not a Moorean fact. “Free will” is a semi-technical term. Free will is something of which we have only a vague idea before we enter into philosophizing about it. Given what I have claimed to be the current state of knowledge in free will research, it is reasonable to conclude that we do not (yet) know we have free will.

Some will not be convinced. Lewis, in a brief argument for compatibilism, suggests this kind of Moorean response:

The best argument for compatibilism is that we know better that we are sometimes free than that we ever escape predetermination; wherefore it may be for all we know that we are free but predetermined. (Lewis 2000, p109)

Note how Lewis's phrasing echoes his above characterization of Moorean facts. Van Inwagen suggests a slightly different argument for the idea that we know we have free will. After claiming that moral responsibility requires free will, he says:

The proposition that often we are morally responsible for what we have done is something that we all know to be true. (1983, p209)

Van Inwagen emphasizes not our (supposed) knowledge that we have free will, but our (supposed) knowledge that we are morally responsible (and our knowledge that moral responsibility requires free will). Similarly, Strawson 1962 may be seen as claiming that the existence of moral responsibility is a Moorean fact. It may be surprising that we can easily know that (A) is false, but this is a bullet that could be worth biting.

The following should give such respondents pause. If we can easily work out the falsity of (A) based on our commonsense knowledge, then were we to find out the truth of one of the conjuncts, we then could easily work out the falsity of the other conjunct. But this does not seem true. Finding out that incompatibilism is true would not make it easy to work out that action determinism is false. Finding out that action determinism is true would not make it easy to work out that incompatibilism is false. Thus we do not know that (A) is false based on our commonsense knowledge. This is a sketch of a response to the above objection. However, it gives rise to another argument that we don't know we have free will that better avoids the above objection.

3.2. The second argument that we don't know we have free will

The second argument better avoids the above pitfall of the first. The basic form of argument is as follows (where “incompatibilism” and “compatibilism” are read as above):

(4) Either compatibilism is true or incompatibilism is true.

(5) If incompatibilism is true, no one knows there is free will.

(6) If compatibilism is true, no one knows there is free will.

(3) Therefore, no one knows there is free will.

Again the argument is clearly valid. (4) is a logical truth. The success of the argument thus relies on the defense of premises (5) and (6).

3.2.1. The argument for (5)

One might think that (5) is easy to show. Consider, for instance, this argument for it:

(7) Incompatibilism is true. (Assumption)

(8) No one knows that action indeterminism is true.

(9) Therefore, no one knows there is free will. (From (7) and (8))

Once we discharge the assumption in (7), we are left with (5). The problem with the above argument is that (9) simply does not follow from (7) and (8). To see this, consider this parody argument:

(7') The existence of water is incompatible with the non-existence of H₂O. (Assumption)

(8') No one knew (in 1700) that H₂O existed.

(9') Therefore, no one knew (in 1700) that water existed. (from (7') and (8'))

Both (7') and (8') are true, but (9') is clearly false. This style of argument, then, fails.

I suggest instead the following argument for (5):

(10) If incompatibilism is true, it is not the case that, were we to find out that incompatibilism is true, at least one of us would thereby have conclusive evidence that action indeterminism is true.

(11) But if incompatibilism is true *and someone knows there is free will*, were we to find out that incompatibilism is true, at least one of us would thereby have conclusive evidence that action indeterminism is true.

(5) Therefore, if incompatibilism is true, no one knows there is free will.

As I use the term, if someone has conclusive evidence for a proposition they are in a position to easily work out that the proposition is true. Thus if I have conclusive evidence that action indeterminism is true, I am in a position to easily work out that it is true.

(11) should be relatively uncontroversial. If incompatibilism is true, and someone knows there is free will, were we all to find out that incompatibilism is true, she would retain her knowledge that there is free will. After all, why would finding out some further truth about free will take away her knowledge that there is free will? I, at least, can see no reason for thinking it would. But if, after coming to know that incompatibilism is true, she would know both that there is free will and that incompatibilism is true, she would then possess conclusive evidence that action indeterminism is true. That there is free will and that free will requires the truth of action

indeterminism quite obviously together entail that action indeterminism is true. Thus, if incompatibilism is true, were we to find out it is, anyone who knows there is free will would possess conclusive evidence that action indeterminism is true. That is, (11) is true.

(10) states that, if incompatibilism is true, it's not the case that, were we to find out it is, someone would thereby have conclusive evidence that action indeterminism is true. The basic idea behind (10) is that finding out that incompatibilism is true would mainly involve various types of a priori philosophical reasoning, and no substantive empirical research in physics. Gaining conclusive evidence that action indeterminism is the case, however, would involve (indeed, could only involve) almost exclusively such research. Thus our finding out that incompatibilism is true would not give us conclusive evidence that action indeterminism is true.

The above argument for (10) may be stated as follows:

(12) If incompatibilism is true, then it is not the case that, were we to find out it is true, we would thereby do substantive empirical research in physics.

(13) It is not possible for us to gain conclusive evidence that action indeterminism is true without us doing such research (some of which has not yet been done).

(10) Therefore, if incompatibilism is true, it is not the case that, were we to find out that incompatibilism is true, at least one of us would thereby have conclusive evidence that action indeterminism is true.

I take (13) to be obvious. We simply cannot figure out that some of our actions (or their immediate causal antecedents) are undetermined from the armchair, even reasoning from the

knowledge we currently possess. Instead, we would have to do sophisticated work in physics to even have a hope of finding out that action indeterminism is true.

(12) is more controversial. Some might think that, were we to find out that incompatibilism is true, we would thereby do substantive empirical research in physics, and thus also come by conclusive evidence that action indeterminism is true. This is an interesting claim as it is usually thought that whether or not free will is incompatible with determinism is not an empirical, but an a priori, matter. Even if we *would* do physics in discovering the truth of incompatibilism, we *needn't*. This suggests the following modification of (10):

(10') If incompatibilism is true, it is not the case that, were we to find out that incompatibilism is true (in a way that does not involve substantive empirical research in physics), at least one of us would thereby have conclusive evidence that action indeterminism is true.

This modified premise calls for a similar modification to (11):

(11') If incompatibilism is true and someone knows there is free will, were we to find out that incompatibilism is true (in a way that does not involve substantive empirical research in physics), at least one of us would thereby have conclusive evidence that action indeterminism is true.

The case for (11') remains untouched. The case for (10') is simply that (13) is true. If one were to find out that incompatibilism is the case without doing empirical work in physics, one would not thereby come to have conclusive evidence that action indeterminism is true precisely because one cannot get such evidence without doing such empirical work. Thus if one doubts (10),

though I do not, we may instead appeal to (10'). If incompatibilism is true, no one knows there is free will.^{vi}

3.2.2. *The argument for (6)*

The argument that if compatibilism is true no one knows there is free will runs as follows:

(17) No one knows (or could easily work out) that compatibilism is true.

(18) Even if compatibilism is true (but no one knows it is), it is not the case that, were we to find out that action determinism is true, someone would thereby have conclusive evidence that compatibilism is true.

(19) But if compatibilism is true *and someone knows that there is free will*, then were we to find out that action determinism is true, someone would thereby have conclusive evidence that compatibilism is true.

(6) Therefore, if compatibilism is true, no one knows that there is free will.

The argument for (19) mirrors our argument for (11) (and (11')) above. If compatibilism is true, and someone knows there is free will, were we to find out that action determinism is true, she would retain her knowledge that there is free will. After all, why would finding out some scientific fact that is perfectly compatible with free will take away her knowledge that there is free will? But if, after coming to know that action determinism is true, she would know both that there is free will and that action determinism is true, she would then possess conclusive evidence that compatibilism is true. That there is free will and that action determinism is true quite obviously together entail that compatibilism is true. Thus, if compatibilism is true, were we to

find out that action determinism is true, anyone who knows there is free will would possess conclusive evidence that compatibilism is true. That is, (19) is true.

The argument for (17) I have set out in detail in section 3.1.2 above. The basic idea is that the arguments for incompatibilism are good and have not been conclusively rebutted. I am a compatibilist, and have given various responses to incompatibilist arguments (Kearns 2008, 2011, 2012). Still, do we *know* that compatibilism is true? Given the state of the current debate, I doubt it. Some compatibilists would not. Daniel Dennett (whose 2003 is perhaps most emphatic about this) may think not only that free will is compatible with determinism, but that some of us *know* it is. Such a compatibilist will not be convinced by the above argument. Still, (17) is the default view. Compatibilism is a controversial philosophical position. One would need a compelling argument for the idea that we know it is true. Furthermore, even if some of us do know compatibilism is true, it is likely that most of us do not (certainly those who have not done philosophy do not). We may thus present an argument similar to that above for Very Weak FWA:

Very Weak FWA: Almost no one knows whether or not normally functioning adult humans have free will.

(I take it the modifications to the argument are obvious enough.) Lastly, even those people who know that compatibilism is true (if there are any) cannot conclude that they know there is free will because, as we shall see, there are other arguments for Weak FWA.

The basic case for (18) is that, if compatibilism is true, it would take more for us to know it is (given our current state of knowledge) than finding out that action determinism is true. Such a scientific discovery cannot establish the truth of this controversial philosophical thesis. The force

of the arguments for incompatibilism is independent of the truth or falsity of action determinism and thus these arguments retain their persuasiveness even were we to find out that action determinism is the case.

We may put the argument for (18) like this:

(20) If compatibilism is true (but no one knows it is), it is not the case that were we to find out that action determinism is true, we would thereby do substantive philosophical work.

(21) It is not possible for us to gain conclusive evidence that compatibilism is true without doing substantive philosophical work (some of which has not yet been done).

(18) Therefore, if compatibilism is true (but no one knows it is), it is not the case that, were we to find out that action determinism is true, someone would thereby have conclusive evidence that compatibilism is true.

This argument clearly parallels the argument for (10). (20) expresses the idea that action determinism's being true is a scientific, not a philosophical, matter. Philosophers would have little to contribute to the discovery that all our actions (and their immediate causal antecedents) are determined.^{vii}

(21) captures the thought that discovering the truth of compatibilism is in large part a philosophical endeavor. Indeed, even if its discovery did involve empirical work (perhaps including the collection of folk intuitions), it would not involve work in physics. To deny (21) would be to think it possible to work out, given our current stock of knowledge, that

compatibilism is true, without doing any philosophy. No positive philosophical arguments need be put forward, no rebuttals of incompatibilist arguments, etc.

Though I think (21) is very plausible, it is, to my mind, the weakest link in this argument for Weak FWA. Some philosophical positions *have* arguably been refuted through empirical means. For example, presentism (the view, roughly, that only the present is real) is widely considered to have been refuted by the theory of relativity, according to which there is no privileged absolute present. It could be argued, then, that incompatibilism might similarly be refuted (and compatibilism established) by physics.

There is a difference between these cases. Action determinism is not formally inconsistent with incompatibilism. Of course, the *conjunction* of action determinism and the claim that we have free will does formally entail that incompatibilism is false. And if we know we have free will, we presumably know it (at least in part) empirically (via our sense of freedom plus our *a posteriori* knowledge of the practices of moral responsibility). Thus one could argue that, if compatibilism is true, we can refute incompatibilism empirically by coming to find out that action determinism is true. Perhaps this is right. If so, my argument for FWA fails. Still, we are left with another interesting result—incompatibilism is open to empirical refutation. Another interesting result is that, given I have shown that if incompatibilism is true, no one knows there is free will, but I have failed to show (and it is in fact false) that if compatibilism is true, no one knows there is free will, compatibilists have the following argument for compatibilism: we know there is free will, if incompatibilism is true, we don't, therefore compatibilism is true.

As a compatibilist I am attracted to this line of argument, but I have my doubts. If we were to find out that action determinism is true, we certainly would have refuted libertarianism

(incompatibilism plus the thesis that there is free will), but would we have refuted hard determinism (incompatibilism plus the thesis that action determinism is true)? Discovering action determinism is true would transform the philosophical debate into one between soft determinists and hard determinists. Both sides would have strong arguments for their position. It is very plausible (even if it falls short of knowledge) that there is free will. Soft determinism would therefore also become very attractive. (Obviously, compatibilism is attractive for other reasons, and so soft determinism would also be attractive for those reasons.) Similarly, however, hard determinists could (and surely would) appeal to the very plausible incompatibilist arguments described above to back up their position. Of course, these arguments might be unsound, but we would not know they were even if we knew that action determinism is true. Whether free will is compatible with action determinism would still be a philosophical matter. Thus whether we *have* free will would also be a philosophical matter, which in part would rest on the conclusion of the debate between soft and hard determinists. And as long as that debate remained unresolved, we would not know that we have free will.

3.3 We don't know we don't have free will

The claim in this section is relatively uncontroversial, and it will be dealt with swiftly. It is that no one knows that we don't have free will. Those in need of an argument for this, consider first the following claim:

(B) At least one possible compatibilist or libertarian account of free will both captures what free will is, and truly describes many normal human adults.

The argument that we don't know we lack free will proceeds as follows:

(25) If anyone knows there is no free will, then she knows (or could easily work out) that (B) is false.

(26) No one knows (or could easily work out) that (B) is false.

(27) Therefore, no one knows there is no free will.

This argument mirrors the first argument that no one knows there is free will. I expect far less resistance to this argument. (25) is defended just as (1) is. I refer you to that defense.

(26) is compelling. There are many sophisticated accounts of free will presented by both compatibilists and libertarians (e.g. Mele 1995, Fischer and Ravizza 1998, Dennett 2003, Clarke 2006, O'Connor 2002, Kane 1996). We are simply not scientifically or philosophically advanced enough to know that either these accounts (or any other account) fail to capture the nature of free will, or that one such account is true but we do not have free will according to this account.

Furthermore, no one, not even the staunchest skeptic about free will, will claim that our lacking free will is something of which we have commonsense knowledge. We simply do not know there is no free will.

4. Conclusions and Consequences

No one knows whether or not we have free will. That is the conclusion of the preceding section, and the main aim of the paper. In this final section I will briefly examine other arguments for this conclusion, other formulations of FWA, and some possible consequences of FWA. To do justice to all the topics covered in this short section, further work is needed. My intent here is simply to highlight these issues.

4.1. Other arguments for FWA

The above arguments for FWA rely on the classic problem of free will. Because there are other challenges to the existence of free will, there are many arguments for FWA of the same type.

Consider the following propositions:

(C) Agent causation is needed for free will and there is no agent causation.

(D) Divine foreknowledge is incompatible with free will and there is such knowledge.

(E) Physicalism is incompatible with free will and is true.

(F) Bivalence concerning future contingents is incompatible with free will and bivalence holds of future contingents.

(G) Epiphenomenalism about consciousness is incompatible with free will and is true.

(H) Action indeterminism is incompatible with free will and is true.

(I) Situationism is incompatible with free will and is true.

Each of these propositions (and others) entail that there is no free will. If we know there is free will, we could easily work out that each of them is false. If you think it implausible of any of these claims that we could do this, you too should seriously consider FWA. Some such challenges can also be recast in the style of my second argument that no one knows there is free will. The details of how to do so, I leave for the reader.

Another type of argument for FWA consists in attacking our evidence for the existence of free will. What, after all, is our evidence? Here are three options: we feel free; we react in certain ways (showing resentment, gratitude, etc.) to other agents; we have developed plausible theories of free will according to which it is not a difficult thing to obtain. None of these seems sufficient

to conclusively establish that we have free will. But if they aren't, and there is no other compelling evidence for the existence of free will, we do not know there is free will. Such an argument can also be used to counter attempts to respond to the argument set out in section 3 above. That is, once the evidence for free will is set out, it becomes far less clear that the existence of free will is a Moorean commonsense fact.

4.2. Other forms of FWA

I have argued for what I have called Weak FWA, and have mentioned Very Weak FWA as a possible fallback position. Three other interesting forms of FWA are:

Moderate FWA: No one can know whether or not adult humans have free will without some of us undertaking hard philosophical and/or scientific work that has yet to be done.

Strong FWA: It is practically impossible for anyone to know whether or not normally functioning adult humans have free will.

Very Strong FWA: It is (metaphysically) impossible for anyone to know whether or not normally functioning adult humans have free will.

If the arguments in this paper successfully show Weak FWA, it seems they also show Moderate FWA. After all, the basic idea of my argument is that the (epistemic) possibility of both incompatibilism and determinism prevent us from knowing we have free will. In order to surmount such a challenge, substantive philosophical and/or scientific work is needed.

Depending on how hard such work is one may also be tempted by Strong FWA, or even Very Strong FWA. My own (optimistic) opinion is that at most Moderate FWA is true. There may come a time when we are in a position to work out that we do, or that we do not, have free will.

4.3 Consequences

I will end this paper by highlighting, but not exploring, some possible consequences of FWA.

Moral Uncertainty: We often act under uncertainty (of physical facts, economic facts, etc.). In particular, we act under *moral uncertainty*. Moral uncertainty is uncertainty of the moral (or normative) truths. There has been much interesting philosophical discussion of how to act under moral uncertainty (see Lockhart 2000, Ross 2006, Sepielli 2009). FWA may impact such discussion, or make it all the more urgent, because the thesis that we (don't) have free will (specifically, MR-freedom) entails the moral claim that we are (not) morally responsible. Given FWA, we are ignorant of whether people are morally responsible. This raises the question of how we should act in circumstances in which the question of whether someone is morally responsible is pressing. Two phenomena in particular seem relevant: our reactive attitudes, and the practice of punishment and reward. I shall comment briefly on each.

Reactive Attitudes: If no one knows there is free will, then our attitudes of resentment, gratitude, etc. may be ill-placed. Plausibly, it is not enough that someone is morally blameworthy for an action to warrant our resenting her. We must also be in a position to know she is morally blameworthy. In the face of uncertainty about responsibility, we should be cautious in our reactive attitudes.

Punishment and Reward: Similarly, if no one knows there is free will, our practices of punishment and reward may need revision. Our current practices seem at least partially based on the idea that humans enjoy free will and moral responsibility. If we do not in fact know humans are like this, such practices are far less secure.^{viii}

Deliberation: When we deliberate, we arguably assume we have open alternatives, each of which we are able to choose. If FWA is correct, this assumption does not amount to knowledge. Given this, our deliberative practices may involve a degree of irrationality.

The Meaning of Life: Whatever the meaning of life is (that is, whatever it takes for a life to be meaningful) it is very plausible that what meaning our lives have depends on whether or not we have free will (thus, for example, the existentialist idea that the meaning of life is to be authentic to our freedom is correct only if we in fact have such freedom). One immediate consequence of FWA, then, is that we are not currently (and may never be) in a position to know exactly what makes our lives meaningful. Further than this, however, our very ignorance of whether or not we have free will may affect the meaning of our lives. After all, we have already seen that our lack of knowledge about free will may affect the appropriateness of blame and punishment. Thus, for example, consider the idea that an authentic and meaningful existence involves realizing that we have free will (or even that we lack it). If FWA is true such ideas are false.

Free Will Research: If we don't know whether we have free will, but we can work out whether we do, it is very important that we attempt to find out. Discovering we do would help justify current legal and moral practices. Discovering we do not would make apparent how much such practices are in need of revision. If we cannot ever discover the truth about whether we have free will, much philosophy needs to be done examining the consequences of our lack of knowledge. As I have suggested above, FWA itself has significant consequences in many areas of our lives. Philosophers are tasked with elucidating and elaborating these consequences.

This last is a call to arms. Philosophers of free will unite—we have nothing to lose but our ignorance!

References

- Campbell, J K (2007) "Free will and the necessity of the past." *Analysis* 67: 105-111.
- Clarke, R. (2006) *Libertarian Accounts of Free Will* Oxford University Press
- Dennett, D. (2003) *Freedom Evolves* Viking Books
- Fischer J. M. and Ravizza, M. (1998) *Responsibility and Control: A Theory of Moral Responsibility* Cambridge University Press
- Fischer J. M. (2006) *My Way: Essays on Moral Responsibility* Oxford University Press
- Fischer J, Kane R, Pereboom D, Vargas M (2007) *Four Views on Free Will* Blackwell Press
- FitzPatrick, W (2008) "Moral Responsibility and Normative Ignorance: Answering a New Skeptical Challenge" *Ethics* 118: 4: 589-613.
- Guerrero, A (2007) "Don't Know, Don't Kill: Moral Ignorance, Culpability, and Caution" *Philosophical Studies* 136: 59-97.
- Harman, E. (2011), "Does Moral Ignorance Exculpate?" *Ratio*, 24: 443–468.
- Honderich, T. (2002) *How Free Are You?* (Second Edition) Oxford University Press
- Huemer, M. (2000) "van Inwagen's Consequence Argument." *Philosophical Review* 109: 525-544
- Kane, R. (1996) *The Significance of Free Will* Oxford University Press
- Kane, R. (2005) *A Contemporary Introduction to Free Will* Oxford University Press

- Kane, R. (2012) *The Oxford Handbook of Free Will (Second Edition)* Oxford University Press
- Kapitan, T. (2002) "A Master Argument for Incompatibilism?" *The Free Will Handbook*, (ed. Robert Kane) Oxford University Press 127-157.
- Kearns, S. (2008) "Compatibilism Can Resist Prepunishment" *Analysis* 68: 250-253
- Kearns, S. (2011) "Responsibility for Necessities" *Philosophical Studies* 155: 307-324
- Kearns, S. (2012) "Aborting the Zygote Argument" *Philosophical Studies* 160: 379-389
- Lasonen-Aarnio, M. (2008) "Single Premise Deduction and Risk" *Philosophical Studies* 141: 157-173
- Lewis, D. (1981) "Are We Free to Break the Laws?" *Theoria* 47: 113-21
- Lewis, D. (1996) "Elusive Knowledge" *Australasian Journal of Philosophy* 74: 549-567
- Lewis, D. (1993) "Evil for freedom's sake?" *Philosophical Papers* 22: 149-172
- Lewis, D. (2000) *Papers in Ethics and Social Philosophy* Cambridge University Press
- Lockhart, T. (2000) *Moral Uncertainty and its Consequences*, Oxford: Oxford University Press.
- McKenna, M. (2008) "A Hard-Line Reply to Pereboom's Four-Case Manipulation Argument." *Philosophy and Phenomenological Research* 77: 142-159
- Mele, A. (1995) *Autonomous Agents* Oxford University Press
- Moore, G. E. (1939) "Proof of an External World" *Proceedings of the British Academy* 25
- O'Connor, T. (2002) *Persons and Causes* Oxford University Press

- O'Connor, T (2013) "Free Will", The Stanford Encyclopedia of Philosophy (Spring 2013 Edition), Edward N. Zalta (ed.), forthcoming URL = <http://plato.stanford.edu/archives/spr2013/entries/freewill/>.
- Pereboom, D. (2001) *Living Without Free Will* Cambridge University Press
- Pereboom, D (2005) "Defending Hard Incompatibilism," *Midwest Studies* 29: 228-47
- Pereboom, D. (2006) "Kant on Transcendental Freedom," *Philosophy and Phenomenological Research* 73: 537-67
- Pereboom, D. (2008) "A Hard-Line Reply to the Multiple-Case Manipulation Argument," *Philosophy and Phenomenological Research* 77: 160-70
- Rieber, S. (2006) "Free Will and Contextualism" *Philosophical Studies* 129.2 223-252
- Rosen, G (2004) "Skepticism about Moral Responsibility" *Philosophical Perspectives* 18: 295-313
- Ross, J. (2006) 'Rejecting Ethical Deflationism,' *Ethics* 116: 742-68.
- Sepielli, A. (2009) 'What to Do When You Don't Know What to Do,' *Oxford Studies in Metaethics, Vol. IV*, Oxford: Oxford University Press: 5-28.
- Smilansky, S. (2002) *Free Will and Illusion* Oxford University Press
- Strawson, G. (1991) *Freedom and Belief* Oxford Clarendon Press
- Strawson, P. F. (1962) "Freedom and Resentment" *Proceedings of the British Academy* 48:1-25
- Van Inwagen, P. (1983) *An Essay on Free Will* Oxford Clarendon Press

Vargas, M. (2009) "Revisionism about Free Will: A Statement and Defense" *Philosophical Studies* 144: 45-62

Vilhauer, B. (2009a) "Free Will and Reasonable Doubt" *American Philosophical Quarterly* 46: 131-140

Vilhauer, B. (2009b) "Free Will Skepticism and Personhood as a Desert Base" *Canadian Journal of Philosophy*, 39: 489-511

Vilhauer, B. (2012) "Taking Free Will Skepticism Seriously" *The Philosophical Quarterly* 249: 833–852

Vilhauer, B. (2013) "Persons, Punishment, and Free Will Skepticism" *Philosophical Studies* 162: 143-163

Williamson, T. (2002) *Knowledge and its Limits* Oxford University Press

ⁱ Thanks to Al Mele, Randy Clarke, Josh Rasmussen and two anonymous referees for helpful comments.

ⁱⁱ A related position, according to which we rarely know whether an agent is morally responsible on a particular occasion, is defended in Rosen 2004. See Harman 2011, FitzPatrick 2008, and Guerrero 2007 for replies. My argument for FWA is entirely unrelated to Rosen's. Arguably, Kant is a free will agnostic (see Pereboom 2006 for further discussion). It is somewhat difficult to show that FWA is an overlooked position, but one can get a sense that it is by looking at recent discussions of the various positions in the free will debate. FWA is mentioned in none of the articles on free will or moral responsibility in the *Stanford Encyclopedia of Philosophy* (see, for example, O'Connor 2013); nor is it entertained in *Four Views on Free Will*; *The Oxford Handbook of Free Will*, or *A Contemporary Introduction to Free Will*. FWA is simply not mentioned in the free will literature.

ⁱⁱⁱ Of course, given that I will be bringing up various ideas that challenge the existence of free will, contextualists may counter that I am automatically put in a high standards context. Rather than take this as a problem for my project, I take it as a problem for any form of contextualism that implies this. If simply challenging a claim to knowledge automatically puts someone into a high standards context, this threatens to trivialize debates about what we do and do not know.

^{iv} An event, x , is determined just in case something, y , causes x , and y plus the laws of nature and background conditions render x inevitable (that is, x *must* occur if y occurs with those laws and background conditions). An event is undetermined just in case it is not determined.

^v I will admit that there are simpler deductions than those sketched above, such as single-premise deduction. I will further admit that there are cases in which even this kind of deduction may fail to produce knowledge (see Lasonen-Aarnio 2008 for such cases).

^{vi} There is one more objection to consider. The truth of (10) relies on the truth of the following disjunction: either incompatibilism is false, or it is true and it is possible to know it is true. If incompatibilism is false, then the antecedent of (10) is false and thus (10) itself is true. If, on the other hand, incompatibilism is true, then if it is not possible for us to find out it is true, it would be trivially true that, were we to find out that incompatibilism is true, someone would have conclusive evidence that action indeterminism is true (counterfactuals with impossible antecedents are trivially true, at least on standard proposals). The negation of this counterfactual appears in the consequent of (10). Thus if incompatibilism is true, while it is not knowable, then (10) is false. Interestingly, then, one can resist the above argument that if incompatibilism is true no one knows there is free will by holding that incompatibilism is both true and unknowable.

There are two problems with this type of objection to my argument. First, it is not very plausible that incompatibilism is true and unknowable. If incompatibilism is true, then it is almost certainly true for reasons that incompatibilists point out (e.g. action determinism takes away our ability to do otherwise). If such incompatibilists are right, then future philosophical arguments will likely establish the truth of incompatibilism. If incompatibilism is true, it is knowable.

Second, if incompatibilism is true and unknowable, we may give a different but related argument to that above that we do not know there is free will. Consider:

(14) If incompatibilism is true but unknowable, that we don't know incompatibilism is true is not part of the explanation of why we don't know action indeterminism is true.

(15) If we know there is free will, then, if incompatibilism is true but unknowable, that we don't know incompatibilism is true *is* part of the explanation of why we don't know action indeterminism is true.

(16) Therefore, if incompatibilism is true but unknowable, we don't know there is free will.

Why do we not know that action indeterminism is true (or that it is false)? It is because we have yet to find conclusive *scientific* evidence either way. Our ignorance of the truth of *incompatibilism* does nothing to explain why we are ignorant of the truth of action indeterminism. Thus (14) is true. However, if we know we have free will, then (given incompatibilism is true and unknowable) incompatibilism's being unknown *does* partly explain why we don't know why action indeterminism is true. This is because, in the light of our knowledge that there is free will, our (unavoidable) ignorance of incompatibilism prevents us from having conclusive evidence for the truth of action indeterminism. Thus (15) is true. (16) follows. I conclude, then, that if incompatibilism is true, no one knows there is free will.

^{vii} Again, if one has doubts about this (perhaps finding out action determinism is the case would in fact involve substantive philosophical theorizing), it is certainly true that it needn't. We may modify (18) and (19) in the same way as we did (10) and (11) to accommodate this.

^{viii} See Vilhauer 2009a for a related point. Also pertinent is Pereboom 2001. In general, FWA may have similar consequences regarding reactive attitudes, and punishment, as Pereboom's hard incompatibilism. The hard incompatibilist may be taking on too strong a dialectical burden in attempting to show there is no free will. If she can establish simply that we do not know there is, it is likely that many of the moral conclusions she wishes to draw follow. Ben Vilhauer in particular has done much to address the consequences of our not knowing that we have free will. As well the citation above, see also his 2009b, 2012 and 2013.