

Compatibilism, Incompatibilism, and Impossibilism¹

Kadri Vihvelin

Debates that claim to be about the free will/determinism problem often aren't. Incompatibilism is usually understood as the claim that the truth of determinism entails the non-existence of free will: that there is no possible world where determinism is true and someone has free will. Compatibilism is the claim that the truth of determinism is compatible with the existence of free will: that there are possible worlds where determinism is true and someone has free will. So one would expect discussions of the free will/determinism problem to focus on determinism (and related questions about the metaphysics of laws, causation, and counterfactuals) and arguments about the relevance (or lack of relevance) of determinism to free will. But the literature is mostly pre-occupied with other questions.

Perhaps the main reason for this is that incompatibilists and compatibilists tend, for the most part, to be free will believers, and therefore are quite properly concerned with more than just showing that free will is or isn't compatible with determinism. They also want to show that we in fact have (or at least might have) free will and they believe that they can show this only by providing an *analysis* of free will. And of course providing a philosophical analysis of anything is notoriously difficult. And in doing this, the energies of both sides get diverted away from the debate between them and towards a different debate, a debate with someone I will call the impossibilist.

The impossibilist is someone who thinks that it is *metaphysically impossible* for us to have free will, either because she thinks that our concept of free will is incoherent or because she thinks that free will is incompatible with some necessarily true proposition. Neither the compatibilist nor the incompatibilist is an impossibilist (see below, for explanation), but some of the arguments that are presented as arguments for incompatibilism turn out, on closer inspection, to be arguments for impossibilism.

Another reason for the paucity of debate about determinism is that there are other apparent threats to free will which, though logically independent of determinism, tend to be associated with determinism – physicalism and the view that we are part of the natural order of things, subject without exception to the same kind of laws (deterministic or probabilistic) that govern everything else in the universe. Compatibilists typically think of themselves in the business of defending, not just the compatibility of free will with determinism, but also the compatibility of free will with physicalism and naturalism. Sometimes compatibilists assume that an incompatibilist *must* be someone who believes that free will is incompatible with physicalism and naturalism as well as with determinism. This is a mistake, but of course incompatibilists have traditionally embraced dualism and the doctrine of agent-causation (the view that we cause our actions in something like the way that God is supposed to cause things – by being “prime movers unmoved”). And arguments that are supposed to be arguments for incompatibilism often trade on intuitions that concern physicalism or naturalism rather than determinism; for instance, arguments that try to convince us that if determinism were true, we would not

be different, in any relevant way, from *merely physical or merely mechanical* things – wind-up toys, simple robots, and so on.

My concern in this paper is *only* with the free will/determinism problem; that is, only with the debate between the incompatibilist and the compatibilist. I will be defending compatibilism. But before I can do so, it is important to understand exactly what is at stake in this debate.

Defining the Problem

Let's begin with some definitions that are now standard in the literature. Determinism is a contingent and empirical claim: that the total state of the world at any time, together with all the laws, entails a unique future. Indeterminism is the negation of determinism. There is some dispute about the ways in which indeterminism might be true. Most people agree that indeterminism would be true if the fundamental laws turned out to be probabilistic; some people think that this is in fact the case. More controversially, some people think that the laws are somehow limited in scope, so they don't apply to some kinds of things (e.g. the nonphysical minds of human beings) or they don't apply to all of the behaviors of some of the things (e.g. the freely willed actions of human beings). And perhaps there are other ways in which indeterminism might turn out to be true. These distinctions are important for incompatibilists, but do not matter for my purposes.

Since indeterminism is the negation of determinism, and determinism is a contingent thesis, we can divide the set of possible worlds into two non-overlapping subsets: worlds where determinism is true and worlds where indeterminism is true. Let's define the Free Will thesis as the claim that at least one human-like (non-godlike) creature has free will. We won't assume that the Free Will thesis is true, or even that it is possibly true.

We can now explain the difference between impossibilism, incompatibilism, and compatibilism.

The impossibilist says that free will is metaphysically impossible (or, perhaps, that it is metaphysically impossible for any non-godlike creature) and therefore the Free Will thesis is not only false but *necessarily false*. That is, the impossibilist says that the Free Will thesis is false regardless of whether determinism is true or false, regardless of whether physicalism is true or false, and regardless of whether any other *contingent* claim about the world is true or false.

The incompatibilist may be a libertarian, who believes that determinism is in fact false, in the right kind of way, and that we have free will, or she may be a hard determinist, who believes that determinism is in fact true and so we don't have free will. However, she is not an impossibilist because she believes that the truth or falsity of determinism is *relevant* to the question of whether we have free will. She believes that there are possible worlds where human-like creatures have free will; that is, she believes that the Free Will thesis is at least *possibly true*. But she believes that the Free Will thesis and determinism

cannot both be true. She believes that the *only* Free Will worlds are indeterministic worlds.

The compatibilist is someone who believes that the Free Will thesis and determinism can both be true; that is, she believes that the set of Free Will worlds is non-empty and includes deterministic worlds.

Figures 1-3 illustrate the fundamental differences between these three positions. If we understand each figure to circumscribe possible worlds, the crosses and checks represent the defining claims of each position. A cross asserts that the set of worlds is empty, a check that it is non-empty. The unfilled boxes are ones about which the position is neutral.

The impossibilist's claims are represented by Figure 1.

	Determinism	Indeterminism
Free Will	✗	✗

Figure 1: Impossibilism

The incompatibilist's claims are represented by Figure 2.

	Determinism	Indeterminism
Free Will	✗	✓

Figure 2: Incompatibilism

The compatibilist and the incompatibilist disagree with the impossibilist and agree with one another that there are worlds where human-like creatures have free will. They disagree with one another about whether determinism is true at any of those worlds.

The compatibilist's claims are represented by Figure 3.

	Determinism	Indeterminism

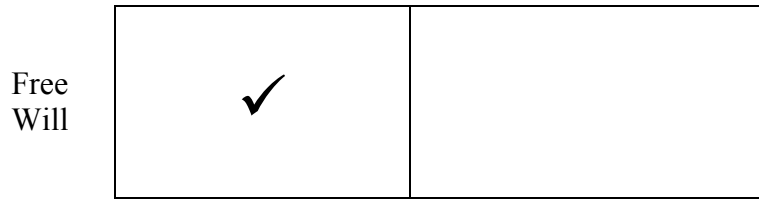


Figure 3: Compatibilism

There have been compatibilists who have claimed more than this. R.E. Hobart famously argued that free will is not only compatible with determinism but positively requires it, at least insofar as our actions are concerned.² But this claim is stronger than one needs to make to be a compatibilist.

The diagrams nicely illustrate how the compatibilist has less of an argumentative burden to bear than the incompatibilist. An incompatibilist need not be a libertarian but the incompatibilist is committed to saying that there are possible worlds where creatures more or less like us have free will. The compatibilist agrees with all that. It is the incompatibilist who must show something else; she must provide some argument that the worlds with free will all lie on the indeterministic side of the line.

Rules of Debate

Remember that there are objections to free will based on contingent claims independent of determinism (physicalism, naturalism) and there are arguments which claim that free will is metaphysically impossible. Because of this, the only way to provide a fully satisfactory defense of free will is to provide a positive account of what free will is. This is what compatibilists have traditionally tried to do. I agree that every compatibilist should have an account of free will (as should every incompatibilist). There is no other way of meeting the challenge of the impossibilist. However, debating the virtues of rival accounts of free will is not what the incompatibilist/compatibilist debate should be about.

The incompatibilist and compatibilist agree that free will is possible. They might even agree about a long list of necessary conditions for having free will. They might agree, for instance, that free will entails the ability to deliberate and make decisions on the basis of reasons, the ability to remember the past and to anticipate the future, the ability to learn from past experience and to use what one has learned as the basis of one's future deliberation and decision-making, the ability to "step back" from one's own character and to ask questions like "Is this the kind of person I really want to be?" And so on. The incompatibilist and compatibilist might even agree that many of the necessary conditions of free will are compatible with determinism. What they disagree about is whether *indeterminism* is a necessary condition. The incompatibilist needs an argument for this, an argument that has the basic structure:

1. Free will entails X.
2. X entails indeterminism.

3. Therefore free will entails indeterminism.

Since the incompatibilist is not an impossibilist, X must be something that is metaphysically possible.

My defense of compatibilism will consist of a critique of the most important and influential arguments for incompatibilism. I will be arguing that the arguments either fail or turn out to be arguments for impossibilism. In order to lay the groundwork, let's begin by looking at one kind of impossibilist argument.

Fatalism

Impossibilism is the claim that there is no metaphysically possible world where any *nongodlike being* has free will. By "nongodlike" I mean someone who is not omniscient, omnipotent, infallible, infinite, the cause of its own existence, and so on.

There are two very different ways of arguing for impossibilism. The first kind of argument claims that our concept of free will is incoherent or entails something that cannot be satisfied by any nongodlike being. We will look at some examples of this kind of argument in the next section. The second kind of impossibilist argument claims that free will is incompatible with some necessarily true proposition or propositions. The fatalist's arguments fall into this category.

The fatalist is someone who argues, on the basis of claims about truth and time, to the conclusion that we don't have free will. What fatalist arguments have in common is a thesis about truth that I will call "Realism about the Future" (**RF**). **RF** says that the future is no less real than the past or present in this respect: there are detailed and specific truths about the future, including truths about our future actions. **RF** says that there are truths about what you will do in the future even if determinism is false and even if there is no way of knowing, ahead of time, what you will do. The fatalist accepts **RF** and argues from **RF** to the conclusion that we have no free will. The fatalist's arguments have nothing to do with the truth or falsity of determinism or any other contingent claim about the world; they are based only on **RF** together with other alleged necessary truths.

Some fatalist arguments are known to be invalid. For instance:

It's either true that I will do X tomorrow or it's true that I won't do X tomorrow. Suppose it's true that I will do X tomorrow. Necessarily, if it's true that I will do X tomorrow, then I will do X tomorrow. (It would be a contradiction if it were true that I will do X tomorrow and I don't do X tomorrow.) Therefore, I *must* do X tomorrow. I have free will only if don't *have* to do what I do; that is, only if I can do otherwise. Therefore, I have no free will.

This argument, sometimes known as “the fatalist fallacy”, makes the mistake of reasoning from the necessity of a conditional to a claim of unconditional necessity. The fatalist’s invalid argument has the form:

1. P
2. Necessarily, if P then Q.
3. Therefore, it’s necessary that Q. (Q must be the case; Q has to be the case; it cannot be otherwise than Q.)

An example:

1. Jones raises his left hand.
2. Necessarily, if Jones raises his left hand then Jones raises his hand.
3. Therefore, it’s necessary that Jones raises his hand. (Jones must raise his hand; Jones has to raise his hand; Jones cannot do otherwise.)

But not all fatalist arguments are invalid. For instance:

1. I have free will only if I can do otherwise.
2. I can do otherwise only if I can do otherwise given all the facts. (That is, only if my doing otherwise is *compossible with all the facts*.)
3. All the facts include facts about the future, including facts about what I will do.
4. Therefore I cannot do anything other than what I will actually do.
5. Therefore I have no free will.

Note that here the fatalist’s conclusion follows, not from **RF** alone (premise 3) but from **RF** together with the fatalist’s claim about what we mean, or should mean, when we say “I can do otherwise”. If we want to deny the fatalist’s conclusion while retaining **RF**, we must reject the fatalist’s claim (premise 2 of the argument) that “I can do X” entails that I can do X, given *all the facts*.

Here’s one more example of a fatalist argument:

1. I have no control over the past.
2. I have no control over the past because the past is “fixed” and “settled” in the following sense: there now exists a set of true propositions that completely describes the past.
3. **RF** is true.
4. Therefore, the future is “fixed” and “settled” in exactly the same sense that the past is: there now exists a set of true propositions that completely describes the future.
5. Therefore I have no control over the future.
6. I have free will only if I have at least some control over the future.
7. Therefore I have no free will.

This argument draws its intuitive appeal from our commonsense way of thinking about truth and time.

We are all familiar with the idea that the past is “fixed” and not in our control. “What’s done can’t be undone”. “There’s no use crying over spilled milk.” By contrast, we think that the future is at least partly “open”, not “fixed”, in our control, and up to us. We believe that we have free will with respect to the future, not with respect to the past.

There is an explanation for this contrast between our beliefs about the past and the future, an explanation that has to do with the fact that the direction of causation always runs from past to future, not future to past. Our choices and actions cause future events; they never cause past events. What we do makes a difference to the future; we cause the future to be what it would not have been had we not done what we did. But what we do makes no difference to the past; we don’t cause the past to be what it would not have been had we acted differently.

The fatalist rejects this explanation. The fatalist says that we have no control over the past *because* there now exists a set of true propositions about everything that happened in the past (premise 2 of the argument). The fatalist thinks that we have the false belief that we have control over the future *because* we reject **RF** and therefore have the false belief that there are no truths about what we will do in the future. But we are wrong, says the fatalist. **RF** is true and it follows that there now exists a set of true propositions that completely describes the future (premise 4). Given this, *and given premise 2*, it follows that I have no control over the future *for the same reason that I have no control over the past*. However, if we reject premise 2, the conclusion does not follow.

Most philosophers think that fatalist arguments are bad arguments because they conflate truth with necessity (either metaphysical necessity or the kind of relativized necessity we express when we say things like “I have no control over the past”). I agree, but this is not my point. Even if the fatalist had a good argument, it would be an argument for impossibilism, not incompatibilism. Neither determinism nor **RF** is part of our commonsense view, and commonsense tends to turn fatalist when forced to take seriously the idea that there are truths about our future actions. But if determinism is true, then there are truths about *all* our future actions. Given this, we need to be on guard, when looking at arguments for incompatibilism, to make sure that they are not fatalist arguments in disguise. And we must be very careful to make sure that the intuitions appealed to are not the same intuitions that support fatalism.

Fatalism is only one way of being an impossibilist. In the next section we will look at an argument that’s usually thought to be an argument for incompatibilism. We will discover that it is in fact an argument for impossibilism.

The Clarence Darrow Argument

“What has this boy to do with it? He was not his own father; he was not his own mother; he was not his own grandparents. All of this was handed to him. He did not surround himself with governesses and wealth. He did not make himself. And yet he is to be compelled to pay.”³

The argument represented by this quote from Clarence Darrow is widely regarded as an argument for incompatibilism. More specifically, it is regarded as an argument for hard determinism – the thesis that determinism is true and *because of this* we lack free will (in the sense necessary to justify blame and punishment). But how are we supposed to understand the argument?

Here’s one way:

1. We have free will only if we make our selves – that is, only if we cause ourselves to be the kind of persons we are.
2. We don’t make our selves.
3. Therefore we don’t have free will.

But if we understand the argument in this way, then the second premise is false, even if determinism is true. We do make our selves, at least in the sense in which we make other things: we plant gardens, cook dinners, build boats, write books, and, over the course of our lives, we re-invent, re-create, and otherwise “make something of ourselves”. We make ourselves by making choices and performing actions which include, among their consequences, changes in our selves. Insofar as we have the ability to make choices, and the ability to predict the consequences of these choices, and the ability to predict how these consequences will affect and change us, we have control over the kind of persons we turn out to be.

Second try:

1. We have free will only if we are *entirely* self-made selves – that is, only if we have *complete control* over the kind of persons we are.
2. We are not entirely self-made selves.
3. Therefore we don’t have free will.

In this reconstruction of Darrow’s argument, the second premise is true. But this is no longer an argument for hard determinism; it is an argument for impossibilism. The truth of the second premise has nothing to do with determinism; no human being (or any nongodlike being) is, or can be, an entirely self-made self. We all have to start from the raw materials given to us by our genes and early childhood environment; we make choices from a range of alternatives often fixed by circumstances outside our control; the causal upshots of our actions are often neither predictable nor in our control. To the extent that we succeed in re-making the self that was “handed” to us, this is only partly due to our efforts and abilities; luck *always* plays a role.⁴

Darrow's argument counts as an argument for hard determinism, as opposed to impossibilism, only if we can find a way to understand his self-making requirement which requires the falsity of determinism while being satisfiable at indeterministic worlds. If determinism is true, then the causes of our actions can always be traced back to earlier events and factors over which we had no control. If determinism is false, on the other hand, then it seems possible that some of our actions (our decisions, choices, and other mental acts) are caused by us, *and by nothing outside us*. This suggests the following way of understanding Darrow's main premise. In order to have the kind of free will that's necessary for moral responsibility, we must have *ultimate* control over our selves in the following sense: we must be the causal initiators (first causes, "Prime Movers Unmoved") of the choices which cause us to be the adult selves we eventually are. On this reading, the argument goes like this:

1. We have free will only if we are ultimately self-made selves – that is, only if *we* have *ultimate control* over the kind of persons we are.
2. We don't have ultimate control over the kind of persons we are.
3. Therefore we don't have free will.

If determinism is true, the second premise is true. So the argument succeeds in showing that the kind of free will specified by premise 1 does not exist at any deterministic world.

To succeed as an argument for incompatibilism (as opposed to impossibilism), however, there must be indeterministic worlds where we (or other nongodlike creatures) have the kind of free will specified by premise 1. But there are no such worlds.

If it were possible to re-make ourselves from scratch, in a way that gives *us* ultimate control over our later selves, it would be by way of reason. Suppose the most favorable scenario for doing this; suppose that we are literally presented with different types of characters, sets of values, and so on, and given a magic pill which will make us into whatever kind of person we want to be. But how do we choose? We might flip a coin, but if we do this, then *we* are not the cause of our new persona, let alone the ultimate cause. If we want to be the cause of our new self, then we must choose on the basis of what we already are -- our reasons, values, principles, together with our ability to deliberate, our ability to critically evaluate our own reasons, and so on. But this counts as ultimate (as opposed to garden-variety) self-making only if *we* caused ourselves to have the reasons (values, etc.) we already have. And this (on pain of infinite regress) is impossible. Our reasons (values, etc.) were ultimately just "handed" to us – it makes no difference whether the handing was by deterministic causation, chancy causation, or whether they popped into existence *ex nihilo*.

Even if determinism is false, we do not and *cannot* make our selves in the way that this reading of Darrow's argument requires -- by causing ourselves to have reasons for *all* our choices and reasons for *all* our reasons.

Historically, there have been two different ways of arguing for incompatibilism. One kind of argument is based on the idea that free will requires that we have ultimate control over

our actions and thereby our selves. I have just argued that this kind of argument is an argument for impossibilism. The other kind of argument is based on the idea that free will requires that we have real (not just epistemic) options, that when we make a choice, we really have a choice, that what we actually do is not the only thing we can do.

There may be a link between the two arguments, insofar as someone might argue that we have ultimate control over our actions and thereby our selves only if we can do otherwise and this entails indeterminism. But we don't need to assume this link. Our ordinary notion of free will includes the idea that we both make and *have* choices, that we have options, that we can do otherwise. If the incompatibilist can show that having options requires indeterminism, this would count as victory for the incompatibilist.

Some would deny this. They would say that the kind of free will that really matters - the kind that's necessary for moral responsibility - doesn't require options or "alternative possibilities", as they are sometimes called in the literature. I think this view mistaken. I am happy to grant that free will requires options. What's at issue is whether options require indeterminism.

Let's look at some arguments for the claim that determinism deprives us of options.

The Forking Paths Argument

1. We have free will only if we can at least sometimes do otherwise.
2. We can do otherwise only if choosing between actions is like choosing between forking paths: that is, only if more than one action is a lawful continuation of the actual past.
3. If determinism is true, then only one action is a lawful continuation of the actual past.
4. Therefore, if determinism is true, we can never do otherwise.
5. Therefore, if determinism is true, we don't have free will.

This argument,⁵ unlike the Clarence Darrow argument, is an argument for incompatibilism (as opposed to impossibilism). The problem is that it's not much of an argument. Premise 2, once it is stripped of the forking paths metaphor, is an *assertion* of the incompatibilist thesis that we can do otherwise only at possible worlds where determinism is false.

Do we have any reasons independent of incompatibilism to accept Premise 2?

It's often claimed that our commonsense beliefs about what we can and cannot do support incompatibilism and that the incompatibilist sense of 'can' is just our ordinary sense and therefore does not need any further support or argument. Let's take a look at this claim.

Suppose that you have the ability to play the piano (you've taken lessons, you know how to play, your fingers are not broken or paralyzed) and you have the opportunity to do so,

and you know you have the opportunity to do so; you are visiting me, and there is a piano in the living room where we are sitting. You have no reason to play the piano, and you don't take yourself to have any reason to play; however, if you had different reasons (if you wanted to show me how a tune goes, for instance, or if I asked you), you would play. Let's stipulate that you are not a victim of brain control, post-hypnotic suggestion, severe depression, or some other pathology that prevents you from forming the intention to play the piano or from acting on your intention. Let's also stipulate that there is no one standing behind the scenes ready to prevent you from playing the piano should you show any signs of wanting, intending, or trying to do so. Suppose in other words, that this is a straight-forward ordinary case in which you fail to play the piano only because you prefer not to do so. This looks like a case where you *can play the piano*, and most people would agree. But Premise 2 says that the facts, as described, do not suffice for the truth of the claim that you can play the piano. You can play the piano only if *more than this is true*; only if your action of playing the piano is a lawful continuation of past history. And the incompatibilist who defends Premise 2 by appealing to our commonsense beliefs must claim that you have this additional belief if you believe that you can play the piano.

I don't think the incompatibilist's claim is very plausible. There is a much simpler explanation of what people believe, when they believe that they can play the piano in a situation like the one described above: They believe something they might express by saying: "I've got the ability to play the piano and there's a piano here and nothing stops me from playing it." Or: "I've got what it takes and the circumstances are right." Or, to put it yet another way, they believe that they have the ability and the opportunity to play the piano.

Of course, it is open to the incompatibilist to argue that this commonsense belief entails the falsity of determinism; that is, to argue that *if determinism is true, then we never have the ability to do anything we fail to do* or to argue that *if determinism is true, then something always prevents us from doing anything we fail to do*. My point is that this alleged entailment is something that needs to be defended by argument; it cannot be simply "read off" our commonsense beliefs.

Is there any other way that the incompatibilist might argue that commonsense supports premise 2? Well, she might appeal to another kind of case, a case that has sometimes been thought to be a paradigmatic example of the exercise of free will. Sometimes we are faced with a decision between possible actions in a case where we have equally strong or perhaps incommensurable reasons for doing each of the alternative actions. Perhaps we have two appealing job offers. Or perhaps we have to decide between a healthy desert and a fattening but delicious dessert. Or perhaps we must decide between doing the right thing and doing what's in our best interest. Suppose your choice is between an apple and chocolate cake; you choose the cake, but believe that you could have chosen the apple instead. What you believe, according to the incompatibilist, is that you could have chosen the apple, given the laws and given all the facts about the past until just before you decided, *including all your reasons and your entire process of deliberation*.

Well, maybe. But I think that even this kind of case is one where we believe something *less* than what the incompatibilist claims. At some time before we make our decision, we believe that we ‘have what it takes’ to decide either way and we believe (if circumstances are favorable – we’ve got enough time to make up our minds, we are not depressed, in a panic, etc.) that nothing prevents us from making either choice. This appears to be neutral with respect to determinism. If we think about what we believe, *after the fact*, it’s even clearer that we are not expressly committed to the belief that determinism is false. If we feel regret or blame ourselves for the choice we actually made, we don’t just think: “I could have chosen the apple instead.” We also think something along the lines of: “If only I had thought about it a bit longer or summoned up a bit more willpower, I would have chosen the apple instead.” That is, even if the actual situation was one where our reasons were so equally balanced that we might as well have flipped a coin, we believe that *if* we had deliberated differently, or longer, we would have discovered stronger reasons for one of our options.

Of course, when we think this, we are also assuming that nothing prevented us from thinking a bit longer or exerting more willpower, and so on. The incompatibilist may argue that this belief entails the falsity of determinism; that is, she may argue that if determinism is true, then something *always* prevents us from doing anything we fail to do. My point, again, is that this alleged entailment is something that needs to be defended by argument.

Let’s take stock. We have been considering the claim that commonsense supports the Forking Paths view of options (Premise 2 of the argument). I have argued that this is far from obvious and that what commonsense believes may, for all we have been shown so far, be compatible with determinism.

What I am not saying: I’m not saying that commonsense believes that determinism is true, or that commonsense believes (even implicitly) that compatibilism is true. I think that commonsense has no opinion, one way or the other, about the truth of determinism, and therefore has never had to confront the question of what would be the case if determinism were true.

Nor am I making a kind of paradigm case argument for compatibilism. I’m not saying that these ordinary cases, where everyone believes that we have free will and options are, by definition, cases where we have free will and options. I am not ruling out, *by stipulation*, the possibility of an argument that will show that incompatibilism is true.

Nor am I claiming that the correct analysis of ‘can do otherwise’ can be extracted from the claims that I have attributed to commonsense, and that this analysis is compatible with determinism. I think that this can in fact be shown, but I do not claim to have shown it here.

What I am saying is that the commonsense view does not clearly and obviously support incompatibilism in the way that is often claimed. The incompatibilist partial definition of ‘can do otherwise’ is global insofar as it says that a necessary condition of a person’s

being able to do otherwise is that there is a possible world with *exactly the same history and exactly the same laws where the person does otherwise*. The commitments of commonsense appear to be more restrictive than this; insofar as commonsense says that something must be held constant in any test of what a person can do, this is something about *the person* and her *surroundings*.

What the incompatibilist needs is an *argument* for the claim that we can do only those things that are a lawful continuation of past history. This brings us to the most important and influential incompatibilist argument in the literature: the Consequence argument.

The Consequence Argument

The Consequence Argument is an argument that is based on the claim that there is some sense in which the laws and the past are necessary, at least relative to us. Different forms of the argument express this necessity in different ways.⁶ Some versions of the argument say that the laws and the past are not up to us; some versions say that the laws and the past are not in our control or in our power; other versions say that they are something that we have no choice about. I will discuss a version of the argument similar to the one discussed by Robert Kane, in this book.

Assume determinism is true. If so:

1. There is nothing I can do to change the remote past or the laws of nature.
2. Necessarily, if the remote past and laws are what they are, my present actions occur.
3. Therefore there is nothing I can do to change the fact that my present actions occur.
4. Therefore I must do what I do; I cannot do otherwise.

This argument has strong intuitive appeal. The first premise seems undeniable. The second premise is entailed by the definition of determinism; it says that my present actions follow from facts about the past together with facts about the laws. But if I can't change the combination of the actual past together with the laws, then surely I can't change what follows from this – my present actions.

But remember the fatalist fallacy:

1. It's true that I will do X tomorrow.
2. Necessarily, if it's true that I will do X tomorrow, I will do X tomorrow.
3. Therefore, it's necessary that I will do X tomorrow.
4. Therefore, I must do X tomorrow; I cannot do otherwise.

The only difference between the Consequence argument and the fatalist fallacy is the claim about the necessity of the past and laws – that is, the claim that we can't change the past or the laws. Without this claim, the Consequence argument would *be* the fatalist fallacy:

1. There are facts about the remote past and the laws of nature.
2. Necessarily, if these facts about the remote past and the laws are what they are, my present actions occur.
3. Therefore, there is nothing I can do to change the fact that my present actions occur.
4. Therefore, I must do what I do; I cannot do otherwise.

So it is very important for the incompatibilist to explain and defend the claim that the laws and the past are necessary; that is, to explain and defend the claim that we cannot change the past or laws.

Note, first, how the necessity of the laws and past cannot be defended. The incompatibilist cannot say that our inability to change the past is due to the fact that the past is already “fixed” or “settled” or consists in “facts carved in stone”. For to say this is to invoke the fatalist’s intuition that something is necessary simply because it is true or a fact. And the incompatibilist cannot say that deterministic laws are unchangeable simply because they entail universal generalizations which are true at all places and all times. Truth at every place and every time is still truth; only the fatalist claims that it is the same as necessity.

We need to be very careful when we talk about our ability or inability to change the facts. Since the Consequence argument is supposed to be an argument for incompatibilism, not impossibilism, we must understand “can change the fact that” in a way that does not make it *metaphysically impossible* to change a fact. Think of a fact about the present that you believe is up to you – a fact about one of your free actions, in a situation where you believe that you can do something else instead. Call this fact – the fact that you do X at time t – ‘F’. Suppose that your beliefs are correct. You live at a world where there is free will (an indeterministic world, if you like) and you do X at time t and you really can do something else instead. Can you change F? Yes and no.

You *cannot* change F in the following way: You cannot change F from what it is originally to what it is after you change it. Suppose that F is the fact that you stay home on June 15, 2006 and suppose that you can go out that day. You *cannot* act so that the following is true: First F is the case (you stay home on June 15, 2006); then F is not the case (you don’t stay home on June 15, 2006). That doesn’t make any sense. If you could change facts in this way, you would be able to do something metaphysically impossible, and no one can do what is metaphysically impossible.

You *can* change F in the following way: There is something that you can do (go out) such that if you were to do it, then it would not be (and would never have been) the case that you stay home on June 15, 2006 and your act would be the event that makes this so. This is the sense in which it is metaphysically possible to change a fact; there is something that you can do such that if you do it, it would not be (and would never have been) the case that F and your act would either be the event that makes it not the case that F or would cause an event that makes it not the case that F. We need a name for this way of

changing the facts; let's call it "causing the facts to be different".

Given this distinction, we can understand the first premise of the Consequence argument as making the claim that we cannot change facts about the past and the laws in this second way; that is, as claiming that there is nothing that we can do that would *cause* facts about the past or the laws to be different. Should we accept this premise?

Let's begin with the past. It seems uncontroversial that we don't have causal power with respect to the past. Backwards causation (e.g. time travel) may be logically possible, but it's not something we are actually able to do. So we should agree that we can neither cause the past to be as it is nor cause the past to be different.

It's less clear what we should say about the laws. On the one hand, it seems that claims about the necessity of the laws need to be defended by defending a particular account of lawhood. If a Humean view of laws as "constant conjunctions" turned out to be correct, would we really be entitled to say that we cannot cause the laws to be different? On the other hand, it seems wildly implausible to think that we can run faster than the speed of light, walk on water, or perform other acts which entail the falsity of the actual laws. And if we were able to perform these law-breaking acts (or perform other acts which cause law-breaking events), then we would be able to cause the laws to be different. So let's agree that we cannot do or cause law-breaking events and for that reason cannot cause the laws to be different.

Now that we have figured out the sense in which we cannot change the past or the laws, we can return to the argument.

1. There is nothing I can do to cause the remote past or the laws to be different.
2. Necessarily, if the remote past and the laws are what they are, my present actions occur.
3. Therefore, I cannot cause my present actions not to occur.
4. Therefore I must do what I do; I cannot do otherwise.

We have agreed that the first premise is true, and the second premise is the definition of determinism. Here, finally, we have an argument that is not an impossibilist argument. But is it valid? If we accept the premises, must we also accept the conclusion?

Let's think it through by using a concrete example. Pretend that determinism is true and I just refrained from raising my hand. This is an ordinary case; no brain control, hypnosis, pathological conditions, and so on. As a compatibilist, I say that I could have done otherwise; I could have raised my hand. But the fact that I did not raise my hand is a logical consequence of facts about the remote past together with the laws. It therefore follows that if I had raised my hand, then either the remote past would have been different or the laws would have been different. The Consequence Argument succeeds if it is *also* true that if I had raised my hand, then my action would have *caused* the remote past or the laws to be different.

Suppose, first, that the laws would have been the same and the past different. Consider one of the possible worlds where this is the case. This is a world where events happened differently in a way that provided me with some reason to raise my hand. Did my hand-raising cause these earlier events? Of course not. The direction of causation was the other way around; the earlier events caused me to have reasons that I did not in fact have, which caused me to raise my hand.

Suppose, second, that most of past history would have been the same, and the laws would have been slightly different. Consider one of the possible worlds where this is the case. This is a world where everything happened exactly as it actually did until shortly before the time when I did not raise my hand, at which point events happened differently in a way that provided me with some reason to raise my hand. The event whereby the history of this world diverged from the history of our world is an event that entails the falsity of *our* laws. Because of this law-breaking event the laws at this world are slightly different from our laws. Did my hand-raising cause this law-breaking event and thereby cause the laws to be different? Of course not. The direction of causation was the other way around; the law-breaking event caused me to have reasons I did not in fact have, which caused me to raise my hand.

If I had raised my hand, either the past or the laws would have been different. But my action would not have *caused* either the past or the laws to be different. And since my action would not have caused the past or the laws to be different, I cannot cause either the past or the laws to be different.

This shows that the Consequence Argument is not valid. We can accept both premises, yet deny the conclusion. Even if determinism is true and it is true that we cannot cause either the past or the laws to be different, it does not follow that we cannot do otherwise.

Conclusion

My defense of compatibilism has been unorthodox. The standard compatibilist defense is by offering an account of free will but I have not attempted to provide even a sketch of an account. I have defended compatibilism by pointing out something that should be obvious but has gone unnoticed in the literature. The incompatibilist is not an impossibilist. Some impossibilists think that our concept of free will is incoherent or self-contradictory or impossible for any nongodlike being to satisfy, and I agree that a fully satisfactory defense of free will should meet these charges by saying enough about what free will is to make it plausible that a human-like being could have free will. The incompatibilist, however, is someone who agrees that free will is possible for human-like beings – but only at indeterministic worlds. Given this, the burden of proof lies with the incompatibilist. Not because the compatibilist claim needs no proof but because the proof is so easily rendered.

To see this, reflect for a moment on the nature of the metaphysical dialectic. To show that something X is possible all that is required is that we describe a possible world where X exists. Not only is this all that is required to prove that X is possible, it is all the proof and

the only kind of proof there can possibly be of the possibility of X. On the other hand, someone who hopes to show that X is impossible must show that there is *no possible world* where X exists, by showing that the description of X entails some logical or metaphysical impossibility.

Accordingly, to show that free will and determinism are compossible we must describe a world at which there is free will and determinism. I have just now— in the preceding sentence — described it. That is all the *positive* argument the compatibilist can give or can be expected to give for her position.

Having given the best and only argument that can be given for compatibilism, it is now the burden of the incompatibilist to demonstrate how this description conceals some logical or metaphysical impossibility. In my view, in the long history of the free will debate no incompatibilist has ever met this burden.⁷

¹ Thanks to John Carroll, Janet Levin, Ted Sider, and Terrance Tomkow for helpful comments and discussion.

² R. E. Hobart, “Free Will as Involving Determination and Inconceivable without It”, *Mind* 63 (1934), pp. 1-27.

³ Clarence Darrow, 1924, “The Plea of Clarence Darrow, in Defense of Richard Loeb and Nathan Leopold, Jr.”, *Philosophical Explorations: Freedom, God, and Goodness*, S. Cahn, ed., New York: Prometheus Books, 1989.

⁴ For discussion of the different ways in which our actions and judgments of moral responsibility are subject to luck, see Thomas Nagel, “Moral Luck”, *Mortal Questions*, Cambridge: Cambridge University Press, 1979, pp.24-38.

⁵ For discussion of versions of this argument, see John Fischer, *The Metaphysics of Free Will*, Blackwell, 1994 and Peter van Inwagen, *Metaphysics*, Westview Press, 2002.

⁶ The classic statement of the Consequence argument is by Peter van Inwagen, *An Essay on Free Will*, Oxford: Clarendon Press, 1983. The secondary literature on this argument is immense; for a summary see Kadri Vihvelin, “Arguments for Incompatibilism”, *Stanford Encyclopedia of Philosophy* (2003), <http://www.plato.stanford.edu> .

⁷ For defense of fatalism, see Richard Taylor, *Metaphysics*, Prentice-Hall, 1992. For defense of impossibilism, see Galen Strawson, *Freedom and Belief*, Oxford: Clarendon Press, 1986. For defense of compatibilism, see Daniel Dennett, *Elbow Room: The Varieties of Free Will Worth Wanting*, Cambridge, Mass.: Bradford Book, 1984; David Lewis, “Are We Free to Break the Laws?” *Theoria* 47 (1981), 113-121; Kadri Vihvelin, “Free Will Demystified: A Dispositional Account”, *Philosophical Topics*, Agency Theory, John Fischer, ed., (2006) forthcoming;. and Susan Wolf, *Freedom Within Reason*, Oxford: Oxford University Press, 1990.
