Journal of Cognition and Neuroethics

How Not To Think about Free Will

Kadri Vihvelin
University of Southern California

Biography

Kadri Vihvelin is Professor of Philosophy at the University of Southern California. Her research focuses on topics in metaphysics and the philosophy of mind and action, including causation, counterfactuals, dispositions, free will, time travel, and mental causation. She is the author of *Causes, Laws, and Free Will: Why Determinism Doesn't Matter*, Oxford University Press, 2013.

Publication Details

Journal of Cognition and Neuroethics (ISSN: 2166-5087). March, 2015. Volume 3, Issue 1.

Citation

Vihvelin, Kadri. 2015. "How Not To Think about Free Will." *Journal of Cognition and Neuroethics* 3 (1): 393–403.

How Not To Think about Free Will

Kadri Vihvelin

Abstract

Our belief that we have free will is one of those entrenched beliefs of commonsense that no one denies until they start doing some philosophy. Some say that our belief that we have free will is incompatible with the existence of truths about our future actions. That's a mistake—and is generally agreed (by philosophers, at least) to be a mistake—the fatalist's mistake. Others say that free will is incompatible with determinism. They say that if determinism turned out to be true, our common sense belief would turn out to be false. Because our common sense belief is so firmly entrenched, some think that we are entitled to conclude that determinism must be false, and must be false in quite specific ways to give us the "elbow room" or the "leeway" or the "robust alternative possibilities" needed for free will. Others think we have no right to reason, from the basis of a commonsense belief to the falsity of something that science tells us (or at least might tell us). But they also assume that the truth of determinism is incompatible with the truth of our common sense belief that we have free will. But is our common sense free will belief really incompatible with determinism? This I take to be the problem of free will and determinism. It's a problem that's been around for quite awhile. I claim, in my book (Causes, Laws, and Free Will: Why Determinism Doesn't Matter, OUP 2013) to have solved it. I'm not going to talk about my solution. What I want to talk about is something that comes under the heading of methodology. I want to talk about how we should think (and talk and write) about free will. But I will begin, as my title suggests, on a more negative note. I offer the following list of don't's for the free will philosopher: don't change the subject, don't do thought experiments, don't rely on intuitions, don't confuse "that's really strange" with "that's impossible," don't worry about hard cases, and don't analyse.

Keywords

Free will, determinism, fatalism, intuitions, thought experiments, Frankfurt examples, Manipulation argument

How Not To Think about Free Will

We've got free will. I'm able to raise my arm—I just did. Now I'm not doing it. But I'm still able to do it. And it isn't just true that I'm able to raise my arm even when I'm not raising it; it's also true that I'm able to choose to raise my arm even when I'm not choosing to do it. And the same goes for lots of other things that we don't do but can do. We are able to do much more than we actually do. We have unexercised abilities, unexercised powers of causing—call them powers of agent-causation, if you want to give them a fancy name. We are able to make choices, on the basis of reasons and reasoning, whether or not we actually do so. We are able to try to do lots of things, whether or not we actually do try. We are able to acquire new beliefs, even if we are too lazy to do the reading or thinking to do so. We are surrounded by unactualized possibilities; we have

abilities we don't exercise (perhaps some abilities we never exercise). We don't have to do what we do. We are able to do otherwise.

I take this to be a fact. Call it 'the Free Will fact.' No one denies it. Well, not quite. It's one of those facts of commonsense that no one denies until they start doing some philosophy.

Some say that the Free Will fact isn't compatible with the existence of truths about the future. They say that if it was "already true" last week—or last month, or last year, or ten billion years ago—that I would fly to Flint and give this talk today, then that's something I had to do. I never had a choice; I was never able to do otherwise. That's a mistake—and is generally agreed (by philosophers, at least) to be a mistake—the fatalist's mistake.

Others say that the Free Will fact is not compatible with determinism. They say that if determinism turned out to be true, this ordinary fact would not be a fact. Some say that we are entitled to conclude that determinism *must be false*, and must be false in quite specific ways to give us the "elbow room" or the "leeway" or the "robust alternative possibilities" needed for free will. Others think we have no right to reason from a commonsense belief to the falsity of something that science tells us (or at least might tell us). But they also assume that the truth of determinism is incompatible with the Free Will fact.

But is the Free Will fact *really* incompatible with determinism? This I take to be the problem of free will and determinism. It's a problem that's been around for quite awhile. I claim, in my book (Vihvelin 2013) to have solved it. (In case it isn't obvious, I am a compatibilist.) I'm not going to talk about that today.

What I want to talk about is something that comes under the heading of methodology—I want to talk about *how* we should think (and talk and write) about free will. But since my time is very short, I will focus on the negative. I will begin by saying how we should *not* think (or talk or write) about free will. Think of these as Rules for the free will philosopher.

First Rule: Stick to the subject—free will.

Don't start talking about something else instead. Don't start talking about moral responsibility. Free will is necessary, but not sufficient, for moral responsibility. Free will is common—wise people have it, foolish people have it, some say that babies and many nonhuman animals have it. Moral responsibility is not so common—no one thinks that babies are morally responsible and there is lots of controversy about when adults are responsible, and more controversy about whether anyone is ever responsible, or whether the concept of moral responsibility even makes sense. But most of this controversy has

nothing to do with free will. We might agree that everyone in this room has free will and would have free will even if determinism turned out to be true. It would not follow that any of us is morally responsible or even that it is possible for anyone to ever be morally responsible. So let's keep this firmly in mind, when we talk about free will, and not slip into those dangerous phrases like "moral freedom," or "the free will that grounds (or justifies or suffices for) moral responsibility."

Or, at least, let's avoid talking in these ways if we hope to make any progress in figuring out what to say about the problem of free will and determinism.

Second Rule: Avoid thought experiments.

Don't get me wrong. Thought experiments are often a useful tool—sometimes a thought experiment is just what's needed to correct a mistake based on failure of imagination.

For instance, if someone says that you must be awake to be morally responsible, then we can show this false by telling a story about a night watchman who falls asleep on the job, so a burglary occurs on his watch. He was asleep when it happened, but he is still responsible because he could and should have been awake. This is a successful thought experiment but note that it has two ingredients—we can all understand what is being described and we all agree about the verdict. It works because it spells out a possibility we had not thought of, or had forgotten about. (It's a counterexample.)

A more complicated example of a good thought experiment is the story that Sydney Shoemaker told to refute the claim that there can be no time without change (Shoemaker 1969). Shoemaker told a story about a possible world in which there are three distinct regions, each of which experiences a local freeze (a yearlong period when there is no change) at regular intervals. And because the freezes happen in a regular pattern for the entire history of that universe, there is good inductive evidence that every sixty years there will be a global freeze. This is a good thought experiment because it is perfectly clear what is being described, and because the story makes us aware of a possibility that, until Shoemaker described it, had not occurred to us.

Unfortunately the free will literature is filled with example of bad thought experiments.

Manipulation Arguments are bad thought experiments. These are stories in which we are invited, say, to imagine people who are "just like us" except that everything they think and do is "remote controlled" by evil neuroscientists. They come in different varieties,[1] but

1. For one well-known example, see Derk Pereboom's description of Professor Plum in Pereboom 2001, 112–113.

they all suffer from the problem of under-description. It is not at all clear what we are being asked to imagine. And they suffer from the problem that people disagree about the verdict—they haven't changed the minds of any compatibilists. So what exactly is the point? Perhaps to make vivid to the uninitiated what a deterministic universe is like? But we don't need a story about manipulation to do that. Whenever I teach my semester long course on Free Will and Determinism, I succeed in depressing my students for several weeks when I tell them what the thesis of determinism is and gradually convince them that they can't just dismiss it, that its truth—or something close enough—is a live possibility. Why are they depressed? Because they think that determinism rules out free will. That, I believe, is a mistake; others disagree. But that's what we should be talking about—whether determinism *really* has this bad consequence.

Frankfurt's alleged counterexample to the Principle of Alternate Possibilities is another example (Frankfurt 1969). In Frankfurt's story there is a mysterious character who, we are told, can prevent you from doing or deciding *anything* you might do or decide. But, in fact, he doesn't interfere with you because he happens to approve of what you do. In that case Frankfurt thought you must still be responsible because no one interfered with your doing what you wanted, even if you couldn't have done otherwise.

Frankfurt's story was supposed to undercut the traditional debate about whether determinism robs us of free will by convincing us that the ability to do otherwise isn't, after all, necessary for moral responsibility. His story didn't work, so his friends and supporters told other stories and an entire literature of "Frankfurt style-examples" sprung up and has lasted more than 40 years, with no signs of stopping or even slowing down. I have argued in print, more than once, (Vihvelin 2000 and Vihvelin 2008) that the stories don't work, that they are underdescribed thought experiments and that when you look more closely at the details, the subject of the stories never loses the ability to do otherwise. No one has said what's wrong with my argument. The response is always: "But, wait, here's another story."

However, it doesn't really matter whether I am right or wrong. The point is simpler. Frankfurt stories fail the two requirements on being a good thought experiment: that we all know what is being described and we agree about the verdict. A counterexample either works or it doesn't. If you have to spend 40 years arguing about whether a counterexample works, your thought experiment is a failure.

You might think, at this point, that this is just a problem for "Armchair Philosophy," and that the fix is to leave the armchair and to run some actual experiments.

But the questions that are used by the Experimental Philosophers include their own thought experiments, which are no less murky.

Here is a typical question (Nichols and Knobe 2007).

Imagine a universe (Universe A) in which everything that happens is **completely caused** by whatever happened before it. This is true from the very beginning of the universe, so what happened in the beginning of the universe caused what happened next, and so on right up until the present. For example one day John decided to have French Fries at lunch. Like everything else, this decision was completely caused by what happened before it. So, **if everything in this universe was exactly the same up until John made his decision,** then it ***had to happen*** that John would decide to have French Fries.

Now imagine a universe (Universe B) in which *almost* everything that happens is **completely caused** by whatever happened before it. The one exception is human decision making. For example, one day Mary decided to have French Fries at lunch. Since a person's decision in this universe is not completely caused by what happened before it, **even if everything in the universe was exactly the same up until Mary made her decision**, it **did not have to happen** that Mary would decide to have French Fries. She ***could have decided to have something different***.

The key difference, then, is that in Universe A every decision is **completely caused** by what happened before the decision—**given the past, each decision *has to happen* the way that it does**. By contrast, in Universe B, decisions are not completely caused by the past, and each human decision ***does not have to happen*** **the way that it does**.

Which of these universes do you think is most like ours? (circle one)

What is the question that is being asked? Do all the phrases used—"completely caused," "had to happen, given the past," and "could not have decided anything different"—mean the same thing? Do the people reading this questionnaire understand what these things mean? Are they all thinking about the same thing when they answer these questions? Are some of them perhaps thinking—as the move from "had to happen, given the past" to 'had to happen" and "could not have decided otherwise" suggests—of a world where there is no free will?

In the absence of any answers to these questions, these supposedly scientific thought experiments are no better than the Armchair variety.

Again, this isn't an argument against thought experiments in general. Nor is it necessarily an argument against thought experiments about free will. But the Rule, for free will philosophers, is this: Unless you know exactly what you are doing, and are sure you can do it well, avoid thought experiments (and avoid experimental philosophy).

Third Rule: Intuitions. Avoid them!

Or if you find that you cannot avoid them because all around you philosophers are appealing to intuitions, and arguing on the basis of their intuitions, and urging you to share their intuitions, constructing thought experiments designed to get you to have more intuitions, or writing up questionnaires for non-philosophers so we can have "data" about intuitions that are not confined to the intuitions of an elite group of philosophers, then I say "just don't do it."

Why not?

Because intuitions have no special evidential status, *qua intuition*. Why would anyone think they do? And how did all this talk of philosophical intuitions get going in the first place? This is relatively recent.

Intuitions are just a kind of belief and we don't think that beliefs *per se* have any special evidential status.

Again, I'm not saying it's *always* wrong to appeal to intuitions. Some kinds of belief are more epistemically trustworthy than others, and this may be true of some intuitions as well. Some people are very good at judging what other people are thinking and feeling simply by reading their faces and body language. Others—the autistic and aspergerish—are not so good. These kinds of intuitions don't have any philosophical payoff. But perhaps there are categories of philosophically relevant intuition which are highly reliable. One possible example might be beliefs about causation in particular cases. We have lots of daily experience of causation so *maybe* our intuitions about causation are a trustworthy source of data to constrain our philosophical theorizing.

But free will intuitions are very different from intuitions about causation.

In the case of causation, we have daily experience of particular cases that count as causation and cases that don't. We can tell the difference between one thing following another by co-incidence, and the first thing causing the second. In the case of free will, however, the clear contrast cases are few and far between. We have free will; rocks and plants don't. We are able to make choices we don't actually make and to do things we don't actually do. But beyond these clear starting points, things get confused and unclear very quickly. We all have free will—at least everyone in this room does. But when did we acquire it? At birth? When we learned to crawl, to talk, to ask questions, to argue with our parents? Or, as some of my students tell me, when we left home to go to university? Do we have free will all the time, or only some of the time? Do we have free will when we are asleep? Under the influence of alcohol or drugs? When we are in a state of panic or severely depressed? Do cats have free will? Might some form of artificial intelligence have free will? When I ask my students these questions, they tell me that they have never

thought about these things before, and many of them change their minds about the answers over the course of the semester.

When it comes to questions about free will and determinism, we have a positive reason to distrust our intuitions. Here's why. It's well known, in philosophy, that the fatalist is confused. Truth isn't the same as necessity, of any kind. The fact that there are truths about my future choices and actions does not affect my freedom *in any way*. But many years of trying to explain to my students why the fatalist is confused has convinced me that fatalist thinking runs deep. Some students get it; others never do. And it turns out that there are arguments for fatalism that are mistaken in ways that are much more subtle than the fatalist is usually given credit for.[2] So the situation is this: Even though it's a mistake, many people have the intuition that if it is "already true" what our future will be, then our future is not up to us; they think that truth alone—*regardless of determinism*—would rob us of free will. But if determinism is true, then there are detailed and specific truths about *all* our future choices and actions. So the intuition that determinism robs us of free will should not be trusted, for it might be a fatalist intuition in disguise.

Fourth Rule: Don't confuse "that's really strange" with that's impossible.

Compare for a moment, a very different literature—the literature about the possibility of time travel. Everyone in that literature understands that those who argue that time travel is impossible must show that the supposition that it is possible gives rise to *actual* contradictions (Lewis 1976). It is not enough to say—indeed, we can all agree—that a world where time travel takes place would be a most strange one.[3]

In the free will literature, by contrast, one often hears remarks to the effect that a deterministic world is a very strange one, and we would have to believe very strange—surprising!—things if we combine a belief that determinism is true with a belief that we have free will. For instance, we would have to believe that if I were to raise my arm just now, then either the remote past or the laws would be different.

But sometimes the surprising is true. This is what the history of science teaches us, and if philosophy is to make progress it should sometimes be what philosophy teaches us.

2. I argue this in Chapter 2 of Vihvelin 2013.

3. For argument that time travel is even stranger than Lewis thinks, see Vihvelin 1996.

Fifth Rule: Don't start with the hard cases.

Don't start with the cases where it isn't clear what to say because we don't know enough to know what to say or because we are confused or conflicted about what to say. The free will/determinism problem is the problem of deciding whether the truth of determinism would have the consequence that the Free Will fact is *never* a fact, not even in the easiest cases, the ones about which everyone agrees.

Sixth Rule: Don't analyze.

At least not at the start, not when you are defending the claim that we have free will (against someone who claims we never have it, or against someone who claims that having it is incompatible with determinism). If you proceed in this way, you are opening the door to the counterexample strategy. You are taking on a greater burden than you need to bear—the burden of defending the claim that your analysis gives the "correct" verdict in the hard cases as well as the easy ones.

Compare: You don't need an analysis of 'chair' or 'game' to be entitled to say that there really are chairs and games, nor do you need an analysis to have the right to say that the existence of chairs and games is compatible with determinism. Nor are you thereby committed to the claim that chairs and games are primitive components of reality.

Concluding Remarks

Back to the Free Will fact and the two objections that I mentioned at the beginning—the fatalist's objection and the incompatibilist's objection.

These objections are treated very differently in the current literature. It is almost universally assumed that the fatalist conclusion is wrong and that the only philosophical problem is to show what is wrong with the fatalist's arguments. (Not all of them are as obviously fallacious as the Fatalist Fallacy.)

But no one thinks this way about the hard determinist or the incompatibilist.

I blame this on the fact that argument by poorly described thought experiments and appeals to intuition is now widespread and common.

It wasn't always so. Back in 1983, back in the days when the incompatibilist was accused of making the kinds of mistakes the fatalist makes—of confusing causation with compulsion, descriptive with prescriptive laws, truth with necessity—Peter van Inwagen wrote an entire book (van Inwagen 1983) arguing that he, at least, was not guilty of any such simple mistakes. He claimed that there is an intuitively appealing and not obviously fallacious, argument for incompatibilism. He called it the Consequence argument.

I agree that this argument is not obviously fallacious. I also agree that the disagreement between us is not a merely verbal dispute. He asserts what I deny—that if determinism were true, then the Free Will fact would not be a fact. But, I claim, he is wrong. The Consequence argument fails. So far as *this* argument is concerned, it may be true that determinism is true and we have free will. And though I have devoted some time to this study, I know of no arguments that work.

So the state of play, at the present time, is that we have no reason to believe that the truth of determinism is incompatible with the Free Will fact. In the absence of other reasons—in the absence of some other *argument*—we are entitled to believe that the Free Will fact is a fact, and would be a fact even if determinism turned out to be true.

But I have digressed. I said that I would talk only about methodology, but I have ended up telling you the punch line of my book. I couldn't resist. But I still have free will. So I will exercise it by stopping.[4]

4. This paper is based on a talk presented at the Free Will Conference at the Center for Cognition and Neuroethics in Flint, Michigan on Oct. 11–12, 2014. Thanks to all who participated for their comments.

References

Frankfurt, Harry. 1969. "Alternate Possibilities and Moral Responsibility." *Journal of Philosophy* 66: 820–839.

Lewis, David. 1976. "The Paradoxes of Time Travel." *American Philosophical Quarterly* 13: 145–152.

Nichols, Shaun and Knobe, Joshua. 2007. "Moral Responsibility and the Cognitive Science of Folk Intuitions." *Noûs* 41 (4): 663–685.

Pereboom, Derk. 2001. *Living Without Free Will*. Cambridge: Cambridge University Press.

Shoemaker, Sydney, 1969. "Time Without Change." *Journal of Philosophy* 66: 363–381.

van Inwagen, Peter. 1983. *An Essay on Free Will*. Oxford: Clarendon Press.

Vihvelin, Kadri. 1996. "What Time Travelers Cannot Do." *Philosophical Studies* 81: 315–330.

Vihvelin, Kadri. 2000. "Freedom, Foreknowledge, and the Principle of Alternate Possibilities." *Canadian Journal of Philosophy* 30: 1–24.

Vihvelin, Kadri. 2008. "Foreknowledge, Frankfurt, and Ability to Do Otherwise: A Reply to Fischer." *Canadian Journal of Philosophy* 38 (3): 343–372.

Vihvelin, Kadri. 2013. *Causes, Laws, and Free Will: Why Determinism Doesn't Matter*. New York: Oxford University Press.