



FREE WILL AND MORAL RESPONSIBILITY: ONCE MORE UNTO THE BREACH

Abstract

This paper is a defense of the thesis,

One is morally responsible for what one has done only if there has been some occasion on which one was able to have acted otherwise than one did.

The defense is based on two premises, the apparently trivial proposition,

If one is deciding between doing X and doing Y, and one ought to do X, then if one asks, ‘Which ought I to do, X or Y?’ the right answer to one’s question is, ‘You ought to do X’.

and the extremely plausible proposition,

If one asks, ‘Which ought I to do, X or Y?’ ‘You ought to do X’ is the right answer to one’s question only if one is able to do X.

Keywords: ‘Free will’ · meaning · definition · moral responsibility · ability to do otherwise · alternative possibilities · fault and blame

I apparently hold an immensely unattractive combination of views in all matters concerning free will and moral responsibility. At least, I suppose they must be, since I seem to be the only one who holds them. Today, I’m going see whether I can make them a bit less unattractive. Nothing I say will be particularly new, I’m afraid.

Let’s begin with free will. Or, rather, let’s begin it with the *words* ‘free will’, the *phrase* ‘free will’. I say, don’t use it. The phrase ‘free will’ is no longer fit for philosophical use, and it’s the fault of philosophers that it isn’t. Every item in the vocabulary of philosophy ought to be either a word or phrase used in its ordinary sense or a well-defined technical term, and ‘free will’ is neither. There is, of course, the problem of what work the word ‘will’ is doing in the phrase ‘free will’, but let that go—what I’m about to say would be, details apart, equally applicable to phrases like ‘free action’ and ‘free decision’.

In ordinary speech, the phrase ‘free will’ has two meanings or senses. I’ll call them the narrow sense and the broad sense. In the narrow sense, the two words occur only in the phrase ‘of one’s own free will’ and in phrases that come from this phrase by substitution of a possessive pronoun for ‘one’s’—‘of my own free will’, ‘of her own free will’, of ‘their own free will’ and so on. This narrow sense implies nothing but lack of coercion. To say that Sally signed the contract of her own free will is to say nothing more than that no one threatened her with untoward consequences if she refused to sign.

The broad sense is really more a lack of sense than a sense. Imagine that some ordinary citizen, Sally again, let us say, is reading the morning paper, and the following headline catches her eye: “Psychologists Claim Experiment Shows Free Will is an Illusion.” If Sally is an at all typical ordinary citizen, whatever reactions she may have to these words, one of them will not be, “I wonder what it is that they think they’ve shown is an illusion. What is this ‘free will’, anyway?” No, she will think she understands the phrase reasonably well. And she certainly won’t believe that the psychologists claim to have shown that human beings never act except under coercion. But she *doesn’t* understand those words. They don’t really mean anything in her mouth and ear. They’re just words that have been taken from some learned sphere into ordinary speech and have shed whatever meaning they had in making the transition. In that respect, they are like ‘evolution’ and ‘energy’ (“Acupuncture attempts to correct disruptions in the flows of energy in the body”), ‘vibration’, ‘entropy’. . . Now on rereading the words I’ve just read about the broad sense a few days after I wrote them, I decided I’d need to add a qualification to it, a need that arose more from the example I used to illustrate my point than it did from anything having to do with the meaning of the words ‘free will’. The example suggests the Libet experiment. And what that famous experiment purports to show, or so its results have been interpreted, is that conscious volition is not among the causal antecedents of our acts. And I suppose that I must admit that the man on the Clapham omnibus will say that people whose actions are independent of their conscious volitions lack free will. I am willing to say, therefore, that the everyday broad sense of free will is not entirely vacuous: there are specifiable conditions necessary for its application, conditions observed by everyone who uses the words. But to concede that isn’t to concede much. If I introduce the term ‘adversarial quasi-inversion’ into our conversation and inform you that adversarial quasi-inversion is an irreflexive relation that, by definition, never holds between an object and a perfect intrinsic duplicate of that object, I still have not conveyed to you anything that could reasonably be called a meaning for ‘adversarial quasi-inversion’.

The technical use of the phrase ‘free will’, to which I now turn, is thus not a sharpening or refinement of a concept expressed by some word or phrase of ordinary language, there being nothing there to sharpen or define. I will speak only of a part of the history of the technical term, the part that began in the 1960s when I first came to philosophy. (The leading edge of

analytical thought about the problem of free will and determinism—I call the problem that because I have nothing else to call it—in those days, can be found in a collection of essays edited by Keith Lehrer and called *Freedom and Determinism*¹.) In those days, students were generally introduced to the technical term ‘free will’ in some such way as this.

To have free will is to do things freely. To say that an agent *S* did such-and-such on a certain occasion freely is to say that *S* did such-and-such on that occasion and could have done otherwise.

A bit later in this essay, I’ll suggest that it’s a bad idea to use ‘could have done otherwise’ in the way the phrase is used in this definition, but what I want to talk about now is the sad story of what became of the definition. And what became of it is that it was that it was improved into uselessness. In my view, it made no sense for anyone to want to improve it, for it was a purely abbreviatory definition—what some would call a purely stipulative definition. Abbreviatory definitions stand in contrast to analytical definitions. Consider these three definitions

S knows that $p =_{df} p$ and *S* believes that p and *S*’s belief that p is justified
 x is an ancestor of $y =_{df} y$ has parents and x belongs to every set to
 which both y and the parents of all its members
 belong.²

x is a fusion of the y s $=_{df}$ A thing overlaps one or more of the y s if and
 only if it overlaps x .

The first is a familiar report of an attempt at an analysis of the concept of knowledge—an unsuccessful attempt, most of us would agree. Unsuccessful though *this* attempt be, providing an analysis of knowledge is a project in which success and failure are possible. The second definition is also of that sort, the product of a successful attempt. The third is *not* a report of an analysis of a concept, and is specifically not a report of an analysis of the concept of the fusion of a plurality of objects. Rather, it essentially enables one to say or write ‘ x is a fusion of the y s’ instead of ‘A thing overlaps one or more of the y s if and only if it overlaps x ’. Too many writers on the problem of you-know-what and determinism have regarded the sixties-style definition

S freely *A-ed* at $t =_{df}$ *S* *A-ed* at t and *S* could have done otherwise

as being like the definition of ‘*S* knows that p ’—as an attempt at an analytical definition of free will. And, since a philosopher is someone who thinks otherwise, generally as being a wrong or failed analytical definition of free will.

1 New York, Random House. 1966

2 $\exists x (x \text{ is a parent of } y) \ \& \ \forall z (z \text{ is a set} \ \& \ y \in z \ \& \ \forall v \forall w (v \in z \ \& \ w \text{ is a parent of } v \rightarrow w \in z) \rightarrow x \in z)$.

Consider, for example, the very first paragraph of Galen Strawson's *Freedom and Belief*³:

Are we free agents? It depends on what you mean by 'free'. In this book, the word 'free' will be used in what I call the ordinary, strong sense of the word, according to which to be a free agent is to be capable of being *truly* or *ultimately* responsible for one's actions.

And Robert Kane has recently contended that there are two conditions that an agent must satisfy to be free. One is that some of the agent's actions satisfy a "could have done otherwise" condition similar to the one that figured in the above definition of 'S freely *A-ed* at *t*'. The other is that the agent be "ultimately responsible" for those actions, which is explained as follows

To be ultimately responsible for an action, an agent must be responsible to some degree for anything that is a sufficient reason (a sufficient condition, cause or motive) for the action occurring.

Kane, moreover, has an argument that is supposed to show why this second clause in the definition of determinism is needed. It is essentially this:

Philosophers have long argued about whether free will is compatible with determinism. And, if to have free will is simply to have been able to do otherwise than one has done, they will never settle this question, for 'able to do otherwise' is a problematical phrase. The debates about whether the ability to do otherwise is compatible with determinism are perennial and irresolvable. Fortunately for those who are interested in the question of the compatibility of free will and determinism, however, there is more to the concept of free will than that . . .⁴

And that is precisely what I want to deny. Or, actually, what is precisely what I want to deny is that there is a concept of free will at all. One can, or so I believe, get causation wrong, or knowledge, or identity across time, but one cannot get free will wrong, for it's not there to be either got wrong or got right. A philosopher who says, "Your definition of free will is wrong. Free agents aren't simply people who sometimes have been able do otherwise than they have done. They're people who . . ." and finishes that definition any way you like, is making no more sense than someone who says, "Your definition of 'realism' is wrong. Realism isn't the thesis that there's a material world external to the mind. Realism is the thesis that there are Platonic universals."

Another consequence of the belief that there's a concept of free will is perhaps more annoying to me than it is interesting to you, to wit, the statement that *An Essay on Free Will* proceeds on the assumption that free

3 Oxford: at the Clarendon Press, 1986. P. 1.

4 These views are expressed in Kane's contribution to Robert Kane and Carolina Sartorio, *Do We Have Free Will?: A Debate* (New York and London: Routledge, 2022). The quoted condition on "ultimate responsibility" occurs on p.11.

will is the ability to do otherwise—which I would compare to the statement, “Your proof proceeds on the assumption that a square root of a number x is a number such that the product of that number with itself is x .”

If you still want to use the phrase ‘free will’, and of course you do, let me ask you this. Think of all the things you want to say using this phrase. Could you say these things without using the phrase? If not, you should be suspicious. One ought to be able to state a philosophical thesis without the use of any particular word or phrase. I could revise *An Essay on Free Will* in such a way that the book contained all the same arguments in defense of all the same positions and yet nowhere contained the word ‘free’ (much less the word ‘will’).

The mistaken belief that the words ‘free will’ have any meaning (other than the “narrow” meaning: to act of one’s own free will is to act when not subject to coercion) as terms of ordinary language would do little harm if philosophers were in general agreement about that meaning. But they are not. One philosopher speaks of ultimate responsibility for one’s actions, another says that to have free will is to have alternative possibilities for action of the sort required for one to be morally responsible for one’s acts, a third of a special kind of causation that proceeds entirely from within the agent. The result is, there is no longer one thing that the words ‘free will’ mean. This was not so in the sixties, the seventies, and at least a goodly portion of the eighties. Philosophers in those days were able all to mean the same thing by ‘free will’ for the same reason that lawyers are able all to mean the same thing by ‘tort’, geologists all to mean the same thing by ‘polymictic rock’, and physicists all to mean the same thing by ‘moment of inertia’.

I could go on for quite a while longer about the words ‘free will’. I could talk about the horrible, confused phrases ‘libertarian free will’ and ‘compatibilist free will’. I could protest the rot that has set in the use of all terms related to ‘free will’—for example, the use of ‘incompatibilism’ to refer the incompatibility of *moral responsibility* and determinism. But I had better pass on to moral responsibility. (No need to talk at any length about ‘determinism’. For present purposes, I need say only that determinism is the thesis that the past and the laws of nature determine a unique future. But perhaps I should say something to straighten out something that I have occasionally got wrong—for I seem sometimes to have identified “determinism” both with ‘the world is deterministic’ and ‘the laws of nature are deterministic’, and those are not the same thing. To say that a world is deterministic is to say that it is the only world with its past and its laws. To say that a set of laws is deterministic is to say that every world whose laws are those laws is a deterministic world. I should have made this distinction explicitly and should never have said, as I sometimes have, used ‘a deterministic world’ and ‘a world with deterministic laws’ interchangeably.)

But back to moral responsibility. I’m of two minds about whether to use the phrase ‘moral responsibility’. The ambiguity of the phrase has occasionally been a source of confusion—the ambiguity, I mean, that is illustrated by the pair of sentences

He is responsible for his daughter's education

He is responsible for his daughter's poor education.

The former means—would in most contexts be understood as saying—that it is his duty, and no one else's, to provide his daughter with an education. The latter means—it would in most contexts be understood as saying—that his daughter is poorly educated and that it's his fault that she's poorly educated. I can remember a distinguished historian of science (and a fanatical enemy of belief in free will) saying "Responsibility isn't something that was conferred on me by a magical power of free will. It's something that was taught to me by my parents". But what his parents taught him was to recognize his obligations, and, perhaps, voluntarily to take on new obligations when it was clear that there were unfulfilled needs that he could help to fulfill. And the responsibility free will is relevant to (if it is relevant to anything having to do with responsibility) is the kind involving fault or blame. I have to admit, however, that the historian's confusion is pretty rare, and there's no great need to guard against it. For the purposes of this essay, at least, I'll use the phrase 'moral responsibility'. I won't use it a lot, but it will occasionally be useful. But, I will insist, what people are morally responsible for is not so much their actions as the consequences of their actions. And I don't see that there's any significant difference between an ascription of moral responsibility and an ascription of fault or blame. Let me illustrate what I mean

Ollie says, "Here's another nice mess you've gotten me into, Stanley." What he says implies that he's in a nice mess and that *it's Stanley's fault that* he's in that mess—or that *Stanley is to blame for* his being in that nice mess. It is important to note that the items we generally blame people for and the items we generally say are someone's fault are something like states of affairs, not actions. If, for example, you have promised to feed my cat while I'm on holiday, and I return home to find that my poor cat has starved to death because you did not keep your promise, I'm unlikely to say that it's your fault that you didn't keep your promise, or that it's your fault that you didn't feed my cat. But it's not at all unlikely that I shall say that it's your fault my cat is dead.

I will now state two important philosophical problems. Taken together, their sum is not all that different from the problem we called the problem of free will and determinism back in the sixties—although neither the word 'free' nor any word derived from it—'freely', 'freedom'—occurs in either statement.

Problem 1

There is an agent who is trying to decide which of two incompatible courses of action, A and B, to pursue. Suppose determinism is true. Is it possible that our agent is both able to pursue course of action A *and* able to pursue course of action B?

Problem 2

Suppose the following statement is true: Anyone who is deliberating between two incompatible courses of action A and B—that is, anyone who is in the position of the agent in Problem 1—is either unable to pursue course A or (inclusive) unable to pursue course B. On that supposition, has anything ever been anyone’s fault? Is anyone ever to blame for anything?

I’ll make a few comments on Problem 1, although my main focus will be on Problem 2.

I might have told the story that figures in Problem 1 not from the perspective of a narrator who speaks of the deliberations of a fictional agent in the present tense, but who is rather looking back on those deliberations (and on the agent’s pursuing the course of action decided upon) from a future perspective. And I might then have posed the problem this way

There was an agent who was trying to decide which of two incompatible courses of action, A and B, to pursue, and eventually chose course A. Suppose determinism is true. Is it possible that our agent was able to pursue course of action B?

This is not greatly different in philosophical import from the corresponding present-tense statement of Problem 2. I mention this matter only because it is important that a statement of the problem in the present-perfect tense not be worded as follows

There was an agent who was trying to decide which of two incompatible courses of action, A and B, to pursue, and eventually chose course A. Suppose determinism is true. Is it possible that our agent could have chosen B?

And that is because ‘could have’ is ambiguous—and so therefore is ‘could not have’. (Ambiguous even setting aside the epistemic sense of ‘could have’—“Who was the department chair the year he was hired?”; “It could have been Sally. It was so long ago I can’t be sure”—which isn’t relevant to our concerns.) I was, I believe, J. L. Austin who first pointed out the ambiguity I’m speaking of.⁵ We can bring it out by considering the sentence ‘You could have ended my career’ in two different conversational surrounds:

At your press conference this morning you spoke very incautiously. You could have ended my career. For about ten awful minutes I thought you had. Fortunately, all the journalists present seem to have been idiots, and they weren’t able to put two and two together. But for God’s sake, think about what you’re saying the next time you talk to the press.

5 See his classic essay “Ifs and Cans” in J. L. Austin *Philosophical Papers* (Oxford: at the Clarendon Press, 1961), pp. 155–180. See particularly p. 163. *Philosophical Papers* was edited by J. O. Urmson and G. J. Warnock.

and

At your press conference this morning you had the opportunity for revenge. You could have ended my career. I expected you to. I expected you to tell the press about my bank account in the Caymans, and I was astonished when you didn't. I want you to know that I'm grateful and that I owe you.

In the first of these two pieces of text, 'You could have' means 'You might have' or 'You came very close to'. In the second, it means 'you were able to' or 'you had it within your power to'. If 'could have' is ambiguous, then so, of course, is 'could have done otherwise'. And the ambiguity of 'could have done otherwise' has led at least some compatibilists to think that when incompatibilists speak of an agent who "could have done otherwise" they are using that phrase to mean "might have done otherwise in exactly the same circumstances." Consider these four sentences:

1. A person is morally responsible for what he has done only if he could have done otherwise.
2. A person is morally responsible for what he has done only if he could have done otherwise in exactly the same circumstances
3. A person is morally responsible for what he has done only if he might have done otherwise in exactly the same circumstances
4. A person is morally responsible for what he has done only if he was able to do otherwise.

Sentence (1) is Harry Frankfurt's statement of his famous "Principle of Alternative Possibilities."⁶ Sentence (2) is, according to Daniel Dennett, a premise of the arguments of the incompatibilists.⁷ But sentence (2) is ambiguous; the principle that—according to Dennett—is a premise of incompatibilism is better expressed by (3). Now incompatibilists may well accept (3), and they pretty much all would have before the publication of Frankfurt's "Alternate Possibilities and Moral Responsibility." But it would have been among their conclusions, not their premises. Rather, before Frankfurt, their premise was, or was something close to, (4).

All this confusion would have been avoided if philosophers writing on the free-will problem had always expressed past-tense ability statements by using 'was able to' rather than 'could have'. (I don't mean to imply that 'was able to' is entirely without ambiguity. But 'was able to' is not nearly so apt to cause confusion as is 'could have'.)

I now turn to the connection between the two main topics of my discussion so far, ability and fault or blame. That is, I turn to Problem 2.

6 Harry Frankfurt, "Alternate Possibilities and Moral Responsibility" *Journal of Philosophy* 66 (1969), 829-839.

7 Daniel C. Dennett, *Elbow Room: The Varieties of Free Will Worth Wanting* (Cambridge MA, MIT Press, 1984), Chapter Six, "Could Have Done Otherwise."

A good case can be made for the thesis that if everyone who is faced with a choice between A and B is either unable to choose A or unable to choose B, then nothing has ever been anyone's fault. And that's really about all one can ask of a philosophical position—that a good case can be made for it.

This seems to be a reasonable principle:

If a certain state of affairs is someone's fault (or 'if someone is to blame for a certain state of affairs'), there is something that person did not do and should have done.

So, for example, my cat starved to death while I was on holiday, and I tell you that it was Henry's fault. You no doubt expect me to go on to say that Henry promised to feed the cat and neglected to do so. And you're probably right—but all manner of things I *might* say are consistent with my statement that the cat's death was Henry's fault. For example, that Henry, a veterinary physiologist, performed an experiment on my cat that rendered it unable to digest food. But all those things all have this in common: they all imply that Henry ought in some way to have acted differently from the way he did. An act of Henry's figuring prominently in a causal chain leading to the cat's death is not sufficient for his being responsible for the cat's death. If, for example, Henry returned from a trip to West Africa carrying a hitherto unknown virus that was harmless to human carriers but which, when transmitted to cats, rendered them unable to digest food, and inadvertently infected my cat, we wouldn't say that the cat's death was Henry's fault.

And it seems reasonable to argue as follows. Suppose I am trying to decide whether to do A or B. If what I should do is A, and if I say to someone, "I'm trying to decide whether to do A or B. Which should I do?", "You should do A" is the right answer to my question. (Although not necessarily the right answer for the person so addressed to give—a person who may have false but amply warranted beliefs about my situation.) I mention this apparently trivial point because I'm sometimes told that the 'ought' of moral advice is not the 'ought' of general moral principles. But if that is true, when Kwame Anthony Appiah, in his "Ethicist" column in *The New York Times Magazine*, tells a correspondent that she should do such-and-such, his being right about that is consistent with there being a correct general moral principle from which it follows that she should *not* do that. (I can't help pointing out that the Categorical Imperative was Kant's answer to the question, *Was soll ich tun?*)

It further seems reasonable to affirm that, if someone asks you for advice, it's a presupposition of that request that the course of action that you recommend be something the recipient of the advice is able to do—a point that in my opinion is valid whether the advice sought is moral advice (by which I mean, "what's the right thing for me to do" advice) or some other kind of advice. Imagine, for example, the following exchange:

Where do you think I should take Alison for dinner?"

“My advice is to take her to Auberge du Vieux Puits in Paris. Best restaurant in the world.”

“Are you kidding? I’m talking about *tonight*—and, anyway, I only have a couple of hundred bucks in my bank account.”

“Oh, I know you aren’t *able* to take her there, but that’s still my advice: take her to Auberge du Vieux Puits.”

Suppose then that it’s Henry’s fault that my cat is dead. If what was said above is correct, it follows that there is Henry did something he shouldn’t have done—or perhaps didn’t do something that he should have done. In either case, that he should have acted otherwise than he did. Let’s suppose he should have fed the cat. That means that if Henry had gone to Sally and said something like, “Sally, I’m facing a moral dilemma. I promised to feed Peter’s cat but I’ve been invited to spend a week with the President at Camp David, and I can’t find someone who’ll take over feeding the cat.” And Sally replies—correctly, since what he should do is feed the cat—, “You should feed the cat.” But that advice makes sense only if Henry is able to feed the cat. Since, in the event, he did not feed the cat, it follows that he was able to have acted otherwise than he did. And, of course, if that reasoning is valid, if he was not able to have acted otherwise than he did, it was not his fault that the cat died.

This highly concrete representation of my argument perhaps fails to display the fact that when I say ‘There is something Henry did that he shouldn’t have done’ the act in question needn’t be at all close in space or time to the events of which it is asserted that they are Henry’s fault. “It” might have been a long series of acts when he was an adolescent, acts that resulted in his growing to be a person who is unable to resist certain kinds of temptation.

We have presented a case for this thesis:

If something that has happened is x ’s fault, then x was (at some point) able to have acted otherwise than he or she did.

And from this it follows that if no one has been able to act otherwise than he or she actually did, then nothing that has happened is anyone’s fault. Or, alternatively, no one is to blame for anything that has ever happened. No one is to blame for the deaths of Tyre Nichols and George Floyd, for the invasion of Ukraine, for the Holocaust, for the Atlantic Slave Trade . . .

We may seem now to be in the territory of the Principle of Alternative Possibilities, but it’s not clear that we are. The principle, we have said, in Frankfurt’s original statement was

A person is morally responsible for what he has done only if he could have [= was able to have] done otherwise.

Is this principle true? Actually, it doesn’t seem very plausible even laying aside counterexamples of the type that Frankfurt employs. Consider, for example, an ambulance driver who is given an address by his dispatcher, and

immediately drives at great speed to 78 Downing Street. When he arrives at that address, he discovers that he's at the wrong place, and is far from where he should be. As a result, the stricken person for whom an ambulance was called dies—and would have lived if the driver had gone to the right address. At the subsequent inquiry, the driver insists vehemently that 78 Downing Street is the address he was given, but all dispatches are recorded, and the recording leaves no doubt in anyone's mind that the address he was given was 778 Dowling Street. In fact, the driver's own recorded voice can be clearly heard repeating after the dispatcher, "778 Dowling Street, got it." The reason for his confusion turns out to be his addiction to methamphetamine. The physician the board of inquiry appointed to test him said, "I examined him ninety minutes after he drove to the wrong address, and found a blood concentration of the drug of 2.5 milligrams per liter. He couldn't possibly have remembered an unfamiliar address correctly for more than two minutes. Frankly, I'm astonished that he was able to keep the ambulance on the road all the way to *any* address. I think this hospital's drug-testing protocols are up for review." It is also established that the driver ingested the drug voluntarily and was familiar with the ways in which it impaired cognitive functioning. If you are comfortable with the idea of an agent's being responsible for his or her actions, I am pretty sure you'll judge that he was responsible for having failed to drive to the address the dispatcher gave him. But if what the doctor said was true, he was almost certainly unable to have done otherwise—unable to have driven to the address the dispatcher gave him.

What this example shows is that the schema

A person is responsible for having A-ed only if he was able to do something other than A

has counterexamples—and writers on the Principle of Alternative Possibilities generally seem to suppose that a counterexample to the schema is a counterexample to the principle. If they are right, the story of the ambulance driver is a counterexample to the Principle of Alternative Possibilities that is not a so-called Frankfurt counterexample. But perhaps one could argue that the principle should be interpreted to mean this

A person is morally responsible for what he has done only if he was able to have done *something* otherwise.

That would save the principle from counterexamples of the "ambulance driver" sort, and the revised principle implies that if the ability to have done otherwise than one has done is incompatible with determinism, then moral responsibility is incompatible with determinism. But, or so I will concede without going into the matter, the revised principle is as susceptible to Frankfurt counterexamples as is the schematic principle.

This discussion of the Principle of Alternative Possibilities has been in aid of the following problem:

The central insight of Frankfurt's essay "Alternate Possibilities and Moral Responsibility" is that one may imagine agents who are clearly morally responsible for various things they have done, and one may go on to imagine adding, so to call them, "offstage counterfactual manipulators," beings who have vast (but unexercised) powers, who do not in any way interact causally with the agents, and whose existence and powers and intensions ensure that the agents were unable to act otherwise than they have done. Can this insight be adapted to refute some premise of the above argument for the conclusion

If no one has been able to act otherwise than he or she actually did, then nothing that has happened is anyone's fault?

It can, of course, only if one can construct a counterexample to some principle presupposed by the argument. And the principle would have to be this—for there is no other candidate:

If "You should do A" is the right answer to the question "Should I do A or should I do B?", then the agent is able to do A.

I will tell two stories.

Story 1

Henry, the fellow who has promised to feed my cat while I'm on holiday, has received an invitation to join the President at Camp David for a week. He asks Kwame Anthony Appiah, "Should I send my regrets to the White House and stay home and feed the cat?" Appiah (rightly) tells him, "You should stay home and feed the cat." Although Henry is able to feed the cat, he decides not to take Appiah's advice, and is off to Camp David, leaving the poor cat to fend for itself.

Story 2

All the events of Story 1 take place, and, in addition the following events: Nellie the Nefarious Neurosurgeon, in connection with an agenda we need not specify, very badly wants Henry to go Camp David. She resolves, therefore, that if he decides not to go, she will (using a technique that she has) manipulate his brain in such a way as to cause him to reverse his decision. But Henry decides to go to Camp David and never internally wavers, and Nellie never interacts causally with him at all.

Adding Nellie (and her powers and her intentions) to Story 1 to produce Story 2 has the consequence that Story 2 differs from Story 1 not only by including more events and an additional character, but by entailing that in Story 2 Henry is not able to stay at home and feed my cat. Does it also have the consequence that

You should stay at home and feed Peter's cat

is not the right answer to the question

Should I stay at home and feed Peter's cat or should I accept the President's invitation and go to Camp David?

It certainly looks to *me* as if it does. I don't see how 'You should stay at home and feed Peter's cat' could be the right answer to that question if Henry is unable to stay at home and feed the cat. For suppose that Appiah knows all about Nellie and her powers and her intentions, although for some reason, he's barred from telling anyone about these things. If Henry asks Appiah what he thinks he should do, how could it be right for Appiah to advise him to do something Appiah knows he's unable to do? (I am not saying that the right advice is "Go to Camp David." Perhaps neither "Go" nor "Stay" is correct. Remember the question Wittgenstein asked Russell—whether he, Wittgenstein, should become a philosopher or an aeronaut. If 'You should become a philosopher' was *not* the right answer to that question, it does not follow that 'You should become an aeronaut' *was* the right answer.)

And if "You should stay at home and feed Peter's cat" is not the right advice for Appiah to give Henry, then it's not the case that he should have stayed at home and fed my cat—for if it were, "You should stay at home and feed Peter's cat" *would* have been the right advice for Appiah to give Henry.

Suppose now that Henry has never been able to act otherwise than he does. Then, by the same reasoning, there is nothing that he did not do and should have done. And then, by the principle

If a certain state of affairs is someone's fault, there is something that person did not do and should have done,

nothing is his fault. And, generalizing this reasoning, if no one has ever been able to act otherwise than he or she does, nothing is or ever was anyone's fault. No one is to blame for anything—no one is or was to blame for anything that happened in the camps in Poland in 1944, no lynching has ever been anyone's fault, no blame accrues to anyone because of the Russian invasion of Ukraine.

The reasoning that led to the conclusion, "If no one has ever been able to act otherwise, than they do, then nothing that has ever happened is anyone's fault" may or may not be correct, but it certainly has this property: it in no way depends on the Principle of Alternative Possibilities and is consistent with its falsity. And that statement is consonant with something I've been saying for many decades: the Principle of Alternative Possibilities is of only marginal relevance to the question whether moral responsibility requires free will (or at any rate to the question whether anything can be anyone's fault if no one is able to do otherwise). An appeal to the Principle is one route from the premise that no one is able to act otherwise to the conclusion that moral responsibility does not exist, but there are others and some at least of those others neither depend on the Principle nor can they be refuted by counterexamples involving offstage counterfactual manipulators.⁸

8 This paper was composed for, and delivered at, a conference on free will and moral responsibility at Florida State University in the spring of 2023. I am grateful to many members of the audience for helpful comments and criticism—and especially so to Alfred Mele.