

PETER VAN INWAGEN

MORAL RESPONSIBILITY, DETERMINISM, AND THE ABILITY
TO DO OTHERWISE

(Received 7 December 1998; accepted 28 April 1999)

ABSTRACT. In his classic paper, "The Principle of Alternate Possibilities," Harry Frankfurt presented counterexamples to the principle named in his title: A person is morally responsible for what he has done only if he could have done otherwise. He went on to argue that the falsity of the Principle of Alternate Possibilities (PAP) implied that the debate between the "compatibilists" and the "incompatibilists" (as regards determinism and the ability to do otherwise) did not have the significance that both parties had attributed to it – since moral responsibility could exist even if no one was able to do otherwise. I have argued that even if PAP is false, there are other principles that imply that moral responsibility entails the ability to do otherwise, and that these principles are immune to "Frankfurt-style" counterexamples. Frankfurt has attempted to show that my arguments for this conclusion fail. This paper is a rejoinder to that reply; I argue that he has failed to show this.

KEY WORDS: compatibilism, determinism, free will, Harry G. Frankfurt, moral responsibility, principle of alternate possibilities

Once upon a time (or so we may suppose: historical accuracy is not my principal object in this introductory paragraph) every philosopher writing about "liberty and necessity" argued either that, because all that happened was necessary, there was no liberty of action or that, because there was liberty of action, not all that happened was necessary. In the course of time, however, there arose a sophisticated school of philosophers (Thomas Hobbes, David Hume, and John Stuart Mill, for example) who pointed out that the two sides in the debate about liberty and necessity shared a premise: that liberty and necessity (or, as we should say today, free will and determinism) were incompatible. And these sophisticated philosophers rejected this premise. Since free will and determinism¹ were compatible, they argued, a determined action could be free; and in fact a determined action *would* be free if it were determined in the right sort of way. (Some members of this school went so far as to contend that an action could be free *only if* it was determined in the right sort of way – and hence only if it was determined.) Subsequently to the rise of the sophisticated school, debates about the problem of free will became

¹ Among the philosophers I have mentioned, only John Stuart Mill would have known the word "determinism," which was coined by Sir William Hamilton.



almost exclusively debates about whether it was the new, sophisticated philosophers or the old, naive philosophers who were right. (After all, the naive have certain advantages over the sophisticated: the Emperor may be wearing no clothes. As Orwell is supposed to have said, “There are some ideas that are so wrong that only a very intelligent person could believe them.”) A wealth of bad philosophical terminology was coined in the course of these debates: “hard determinism,” “soft determinism,” “libertarianism,” “contra-causal freedom,” “freedom in the libertarian sense.” It was not until the 1960s that the terminology of the free-will problem was rationalized. Since that decade, most philosophers discussing the free-will problem have taken something like the following three terms to constitute the “basic” or “primitive” vocabulary of the problem:

The Free-will Thesis: Some human beings have free will; that is, at least some human beings at least sometimes have it within their power to (are able to) act otherwise than they do

Determinism: The past determines a unique future (given the past and the laws of nature, the future is determined in every detail)

Compatibilism: The free-will thesis and determinism are compatible (their joint truth is possible).

(*Indeterminism* and *incompatibilism* are, respectively, the denial of determinism and the denial of compatibilism.) Compatibilism, is, of course, what the debates between the *naïfs* and the sophisticates had been about.² (If there is any need for the older terms, they may be defined as follows: Hard determinism is the conjunction of determinism and incompatibilism – which, of course, jointly entail the denial of the free-will thesis; soft determinism is the conjunction of determinism and the free-will thesis – which, of course, jointly entail compatibilism; libertarianism is the conjunction of the free-will thesis and incompatibilism – which, of course, jointly entail the denial of determinism.)³ The adoption of the newer,

² I believe it was Keith Lehrer who coined the term “compatibilism.”

³ But the need for them is very small, and they are almost always better avoided. As for “contra-causal freedom” and “freedom in the libertarian sense,” they must be consigned to the dustbin of philosophical history. Unfortunately, the latter turns up with some frequency in current writings on free will. To my mind, seeing that the phrase “freedom in the libertarian sense” has no possible use is the *pons asinorum* of the problem of free will, for compatibilists and incompatibilists use “is able to do otherwise” in exactly the same sense. The reader who doubts this should consult my essay “When Is the Will Free?” *Philosophical Perspectives Volume 3: Philosophy of Mind and Action Theory* (1989), pp. 394–422 [reprinted in Timothy O’Connor (ed.), *Agents, Causes, and Events: Essays*

more rational terminology was no doubt progress of a sort, but, at a deeper level than terminology, the character of the free-will debate had not changed for many decades. Books and essays on free will were mainly the work of compatibilists, who had for a long time constituted the orthodox, majority party. Why these books and essays continued to be written was not clear, since they were all more or less the same; each explained wearily – more in sadness than in anger, really – that there never had been a problem of free will and determinism because free will and determinism were compatible – for exactly the reasons that had been given by Hobbes, Hume, and Mill. (That is to say, for exactly the reasons that had been given by Hobbes.) There were, however, a few able and respected philosophers who continued to defend incompatibilism. (C. D. Broad in his Cambridge inaugural lecture; R. M. Chisholm; Carl Ginet. The arguments of these incompatibilists were, of course, technically far superior to arguments of Bishop Bramhall, who had debated what was essentially the compatibilism-incompatibilism issue with Hobbes, but they were not fundamentally different. As one who has made some contribution to getting the argument for the incompatibility of free will and determinism clearly stated, I don't mind admitting that none of us incompatibilists has ever contributed much but increased clarity to the debate: there is essentially *one* argument – a very powerful one – for the incompatibility of free will and determinism, and the major contribution of us incompatibilists to the ongoing discussion of the problem of free will has been to make sure that the force of this argument is appreciated.)

In 1969, however, something new happened. Harry Frankfurt published a remarkable (and now classic) essay in which he denied that the debate between the compatibilists and the incompatibilists had the significance that most members of both parties had attributed to it.⁴ The main interest of the free-will problem, for most philosophers, derived from their belief that moral responsibility was impossible without free will – without the ability to do otherwise. This belief was the main reason most philosophers had for *caring* about free will enough to invest time and ink in a debate about whether anyone had it or what it was compatible with. Frankfurt argued

on Indeterminism and Free Will (New York and Oxford: Oxford University Press, 1995), pp. 219–238)]. This reader should consult particularly the long paragraph that runs from page 402 to page 404 (O'Connor, pp. 222–223) and begins “Before going further . . .” I regard this paragraph as the single most important paragraph I have ever written about the free-will problem. I wish people would pay more attention to it.

⁴ Harry G. Frankfurt, “The Principle of Alternate Possibilities,” *The Journal of Philosophy* LXVI (1969), pp. 829–839. Reprinted in John Martin Fischer (ed.), *Moral Responsibility* (Ithaca: Cornell University Press, 1986), pp. 143–152.

that this belief was false. More exactly, he argued that “The Principle of Alternate Possibilities,”⁵ or PAP as its friends call it, was false:

PAP A person is morally responsible for what he has done only if he could have done otherwise.⁶

This principle is extremely plausible, but counterexamples to it are as many and various (and almost as well known) as Gettier counterexamples in epistemology. Here is one that will do as well as any:

Suppose that Gunnar has shot Ridley and is morally responsible for having done so. (Build into the example whatever you think is needed to make this supposition true.) Now add to the case an offstage “counterfactual manipulator,” Cosser, who *would have* caused Gunnar to shoot Ridley (perhaps by direct manipulation of Gunnar’s brain) if Gunnar *had* shown any hesitation about carrying out his long-standing plan to shoot Ridley. (In saying that Cosser is an “offstage” manipulator, we mean that his existence, powers, and intentions are unknown to Gunnar, and that, unless Cosser were forced to carry out his “contingency plan,” nothing he did would have any effect on Gunnar). But, in the event, Gunnar showed no such hesitancy and, as we have said, went ahead and shot Ridley. Obviously we do not, by adding Cosser’s offstage presence and his unacted-on contingency plan to the example, make our story inconsistent: we do not contradict the supposition with which we began, that Gunnar is morally responsible for having shot Ridley. But just as obviously we *do* change the story in another way: the story now entails that Gunnar was unable *not* to shoot Ridley. Look at the matter this way. Pick some moment, *t*, before Gunnar shot Ridley, a moment at which Gunnar still had plenty of time to change his mind about shooting Ridley, but after Cosser had formed his contingency plan and set up his monitoring devices and instruments of neural manipulation. Various “possible futures” lead away from “the world as it was at *t*.” (Only one if the world is deterministic; it may be, of course, that in accepting the above invitation to “build into the example whatever you think is needed to make this supposition true,” you have built indeterminism into the example.) In *all* these possible futures, Gunnar shot Ridley. In some of them – if determinism is false and there *were* such futures – this was because Cosser caused Gunnar to shoot Ridley. In the others, Gunnar shot Ridley without being caused to do so by Cosser. Nevertheless – in *all* these possible futures, Gunnar shot Ridley. And it is not *only* true that Gunnar shot Ridley in all these possible futures. If determinism is true, in none of them did Gunnar change his mind and decide not to shoot Ridley. Even if the world is deterministic, however, it was still true that (owing to Cosser and his plans and powers) in all the closest futures in which Gunnar changed his mind and decided not to shoot Ridley, he shot Ridley. In other words, the addition of Cosser to the example has the following consequence: the story now entails

⁵ More properly, the Principle of Alternative Possibilities.

⁶ Frankfurt, “The Principle of Alternate Possibilities,” p. 829. In this principle, “could have” must not be taken to mean “might have”; it means “was able to.” The ambiguity of “could have” – the “might have”/“was able to” ambiguity – has caused an immense amount of confusion in discussions of PAP and in discussions of the free will problem in general. Here is a pair of examples (adapted from Austin) that illustrates this ambiguity. “You could have exposed me this morning. For God’s sake, watch what you’re saying when you talk to the press.” “You could have exposed me this morning. I want you to know that I’m grateful you didn’t.”

that Gunnar would have shot Ridley *no matter what choices or decisions he had made*. From this it obviously follows that, at *t* and later,⁷ Gunnar was unable not to shoot Ridley. But then Gunnar is morally responsible for having shot Ridley even though he was unable not to shoot Ridley.

Almost a decade after Frankfurt's essay was published, I wrote an essay⁸ (the body of which is contained in my book *An Essay on Free Will*⁹) in which I argued that, although PAP might be false – for just the reasons that Frankfurt had given –, there were other principles that, in conjunction with incompatibilism, entailed that moral responsibility could not exist in a deterministic world. And I argued that these principles could not be refuted by “Frankfurt-style” counterexamples. (These arguments were in aid of the following conclusion: if Frankfurt is right and PAP is false, it is nevertheless true that the compatibilism/incompatibilism debate has just the significance that philosophers who care about moral responsibility have always supposed it to have.) A few years after that essay was published, Frankfurt published a reply to my arguments for these conclusions.¹⁰ It is that reply that I wish to consider in the present essay. I will examine just one of these “other principles” and what Frankfurt says about it. This is the principle that in earlier writings I have called PPP₂ – but, in the present essay, I will call it simply PPP, since I shall not discuss any principle from which it needs to be distinguished by a subscript. (The three P's stand for “principle of possible prevention.”) Here it is:

PPP A person is morally responsible for a state of affairs only if (that state of affairs obtains and) he could have prevented it from obtaining.

Here is the germ of what Frankfurt has to say that is relevant to my contention that PPP is true, that it cannot be refuted by Frankfurt-style counterexamples, and that, in conjunction with incompatibilism, it entails

⁷ It is consistent with the way we have told the story that there was a time earlier than *t* at which Gunnar was able not to shoot Ridley. It should be obvious that a more elaborate example could have been constructed according to which there was no point in Gunnar's life at which he was able not to shoot Ridley.

⁸ Peter van Inwagen, “Ability and Responsibility,” *The Philosophical Review* LXXXVII (1978), pp. 201–224. Reprinted in Fischer, pp. 153–173.

⁹ Oxford: at the Clarendon Press, 1983. See pp. 161–182.

¹⁰ Harry G. Frankfurt, “What We Are Morally Responsible For,” in Leigh S. Cauman, Isaac Levi, Charles Parsons, and Robert Schwartz (eds.), *How Many Questions? Essays in Honor of Sydney Morgenbesser* (Indianapolis: Hackett, 1982). Reprinted in John Martin Fischer and Mark Ravizza (eds.), *Perspectives on Moral Responsibility* (Ithaca: Cornell University Press, 1993), pp. 286–295.

that moral responsibility cannot exist in a deterministic world. He contends that PPP has “nothing at all to do with free will.” Why not? I am not sure. It seems that Frankfurt intends the following paragraph to supply the answer to this question.

The fact that there are . . . states of affairs that a person cannot bring about plainly does not in itself mean that the person lacks free will. Given that the freedom of a person’s will is essentially a matter of whether it is up to him what he does, it is more a matter of whether it is up to him what bodily movements he makes than of what consequences he can bring about by his movements. Imagine [that an equipment malfunction makes it impossible for a certain man to call the police, despite his freedom to move his body in any way he likes, and imagine that this malfunction] is due to negligence on the part of the telephone company; and imagine that because of this negligence, large numbers of people are unable to do various things. These people may quite properly be resentful. But they will be carrying their resentment too far, and attributing too portentous a role in their lives to the telephone company if they complain that the company has through its negligence diminished the freedom of their wills.¹¹

I am far from sure that I see what the point of this paragraph is supposed to be. For one thing, if its point is to show that PPP has nothing to do with free will, its opening words are puzzling; one would expect it to begin

The fact that there are states of affairs that a person was unable to prevent plainly does not in itself mean that the person lacks free will.

But this is a minor point of exposition. The important point is this: the relevance of PPP to free will could hardly be plainer. More exactly, its relevance to the proposition

In a world without free will (i.e., in a world in which no one is ever able to act otherwise), there is no moral responsibility

could hardly be plainer. Its relevance to this proposition is shown by the fact that this proposition follows from PPP and the two premises

In a world without free will, no one is able to prevent any state of affairs (that does obtain) from obtaining

If there is moral responsibility, someone is morally responsible for some state of affairs.

(And, of course, if there is no moral responsibility in a world without free will, and if free will implies indeterminism, then there is no moral responsibility in a deterministic world.) Both these premises seem to me

¹¹ Fischer and Ravizza, p. 294.

to be obvious truths. But perhaps Frankfurt will deny the first of them. He seems to have very particular ideas about how the phrase “free will” should be used: he wants to connect this phrase very closely with the idea of one’s control over the movements of one’s body. Perhaps his particular ideas about the proper sense of the phrase “free will” will lead him to say that even if one has no control over the movements of one’s body, one may nevertheless be able to prevent certain states of affairs from obtaining. I must say that I don’t see how this could be (unless perhaps the states of affairs one is able to prevent pertain to what would normally be the causal antecedents of the movements of one’s body; but I doubt whether that’s his point). However this may be, let us simply avoid all issues related to the proper use of “free will.” Let us drop the phrase from our discussion. I will now present an argument (in which the words “free will” do not occur) for the conclusion that moral responsibility cannot exist in a deterministic world. In this argument, the variable “*p*” ranges over states of affairs, and the operator “N” is to be understood in the following sense:

$Np =_{df}$ *p* obtains and no one is able or ever was able to prevent *p* (from obtaining).

- (1) If determinism is true and *p* obtains, then Np
- (2) A person is morally responsible for *p* only if (*p* obtains and) he is or was able to prevent *p* (from obtaining)
- (3) If (2) is true and if Np , then no one is morally responsible for *p*

hence,

- (4) If determinism is true and *p* obtains, no one is morally responsible for *p*.

(1) is a fairly standard statement of incompatibilism. Now it may be that if “incompatibilism” is by definition the proposition that determinism and the free-will thesis are incompatible, and if we incompatibilists have somehow been misusing the words “free will,” the thesis expressed by (1) shouldn’t be called “incompatibilism.” Still, this thesis is (more or less) what we who call ourselves incompatibilists have been arguing for under the name “incompatibilism.” It’s the thesis, not the name, that is important to us. If necessary, we’ll find another name for our thesis and ourselves.

Premise (2) is PPP.

Premise (3) is undeniable, for its denial is

p obtains and someone is morally responsible for *p*; if *p* obtains and someone is morally responsible for *p*, he is able or once was able to prevent *p*; no one is able or ever was able to prevent *p*.

The conclusion of the argument is equivalent to, “If determinism is true, no one is morally responsible for any state of affairs that obtains.” And this, I think suffices to show that PPP is “relevant” to the question whether moral responsibility can exist in a deterministic world.

Or so it would seem. One could, however, raise the question whether moral responsibility might exist even if no one was morally responsible for any state of affairs that obtained. (Recall that the second premise of our first argument was “If there is moral responsibility, someone is morally responsible for some state of affairs.”) Let us examine this question that someone might raise. A good way to begin a discussion of this question is to reflect on Frankfurt’s title: “What We Are Morally Responsible For.” What *are* we morally responsible for? If some of the things we are morally responsible for (given that we are morally responsible for anything) are not states of affairs, then moral responsibility might, for all we have said so far, exist in a world in which no one was morally responsible for any state of affairs that obtained in that world. I myself believe that we are responsible for things in other ontological categories than states of affairs that obtain – concrete events, if nothing else.¹² Nevertheless, in my view, we *are* sometimes morally responsible for states of affairs that obtain – but let us abandon this clumsy phrase and say instead that we are sometimes morally responsible for *facts*. (A possible state of affairs that does *not* obtain – there having been a woman President of the United States before the twenty-first century, say – is simply a failed candidate for factuality; it is a thing that would have been a fact if things had gone differently, but happens not to be one.) Here is an indisputable (although, unfortunately, disputed) fact:

Millions of Jews were horribly murdered in the Nazi death camps.

Is someone morally responsible for this fact? Well, of course – but what does it mean to say that someone is morally responsible for this fact? Just this:

¹² See “Ability and Responsibility,” or Chapter V of *An Essay on Free Will*.

That millions of Jews were horribly murdered in the Nazi death camps *is someone's fault*.

Or, alternatively,

That millions of Jews were horribly murdered in the Nazi death camps *is something for which someone can be held morally accountable*.

And who is the “someone” – or, rather, who are the “someones,” for morally responsibility for a fact can be shared, and this is one of those cases? That is to say, who is morally responsible (shares in the moral responsibility) for this fact? Opinions differ. Hitler and Himmler and the inner circle of the Nazi Party, certainly. But certainly others as well, although there is legitimate dispute about where the “list” should be cut off: German industrialists, collaborators in France and Poland and other conquered territories, every citizen of the Third Reich who did not actively resist the Nazis, the Allied High Command . . . the members of all these groups have been said by someone to share in the moral responsibility for the fact that millions of Jews were horribly murdered in the Nazi death camps. But no one says that *no one* is morally responsible for this fact. (Not even “Holocaust deniers”: they deny the *fact*; they don't say that the fact exists and no one is responsible for it.) We all agree, therefore, that some people are morally responsible for some facts. (Even those who, like Clarence Darrow, verbally deny the existence of moral responsibility demonstrate in their behavior that they hold some people morally responsible for some facts.)

But suppose we're all wrong. Suppose that *every* time someone says something like “It's your fault your mother had a miserable old age” or “Alice is to blame for the fact that the Women's Shelter had to close for lack of funds” the speaker is somehow just wrong, mistaken, saying something false. Suppose there *are* no truths of that kind. What would follow? It would follow, I think, that no one was morally responsible for anything – whether it was a fact, a concrete event, or a representative of any other ontological category. Consider, for example, concrete events. It seems evident that if a person is morally responsible for, say, Tom's death, that person most *also* be morally responsible for some fact or facts: for the fact that a bullet entered Tom's brain, for the fact that no one called 911, for the fact that Tom was left alone with his heart medicine out of reach . . . for *some* fact or facts. And this, I think, is true even if (as I believe) no reductive analysis of moral responsibility for concrete events in terms of moral responsibility for facts is possible.

If, therefore, (1) [incompatibilism, or whatever the thesis should be called], and (2) [PPP] are both true, determinism is incompatible with the existence of moral responsibility.

Department of Philosophy
The University of Notre Dame
Notre Dame, IN 46556-5639
USA