MOLINIST FRANKFURT-STYLE COUNTEREXAMPLES AND THE FREE WILL DEFENSE

Michael Bergmann

Harry Frankfurt's well-known counterexample to the principle of alternate possibilities (PAP) has recently come under attack by those who argue that the success of that sort of counterexample depends on the falsity of incompatibilism. In response, I argue that, given one controversial assumption, there are Frankfurt-style counterexamples to PAP that don't take the falsity of incompatibilism for granted. The controversial assumption is the Molinist one that something like middle knowledge is possible. I then show how the falsity of PAP causes some trouble for the Free Will Defense standardly offered in response to atheistic arguments from evil.

Harry Frankfurt has proposed a well-known counterexample to the following principle of alternate possibilities:

> PAP. A person is morally responsible for performing a given act only if she could have acted otherwise.[1]

One thing that makes Frankfurt's proposed counterexample to PAP interesting is that it is supposed to be successful even if the sort of moral responsibility at issue is fairly robust – i.e., of the sort in which an incompatibilist and not merely a compatibilist is interested. Recently, however, Frankfurt's criticism of PAP has come under attack precisely because it (supposedly) *fails* when the focus is full-blooded moral responsibility of the sort that incompatibilists care about.[2] The suggestion is that such counterexamples to PAP are successful only if one assumes the falsity of incompatibilism.

In this paper, I will defend Frankfurt's criticism against this charge. My aim is to design a Frankfurt-style counterexample to PAP that doesn't take for granted the falsity of incompatibilism. According to the sort of incompatibilism on which I'll be focusing, freedom and responsibility are incompatible with causal determinism and crucially involve agent causation.[3] Philosophers who hold such a view – call them 'incompatibilist agency theorists' – typically think of themselves as differing from their compatibilist opponents in this way: they believe in *true* freedom and *genuine* responsibility whereas the compatibilists don't (at least not consistently). Thus, the particular question I want to consider is this: Is there a Frankfurt-style counterexample to PAP that doesn't assume the falsity of the incompati-

bilist agency theorist's position – a counterexample that focuses on their robust notion of moral responsibility?

The main conclusion of the paper, to be defended in section 2, is that, given one controversial assumption, there *is* a Frankfurt-style counterexample to PAP that doesn't assume the falsity of incompatibilism.[4] The controversial assumption is something very much like the Molinist position that middle knowledge is possible (i.e., that there are true would-counterfactuals of genuine freedom and that these can be known). A secondary conclusion, presented in section 3, is that my main conclusion creates a difficulty for the Free Will Defense standardly offered in response to atheistic arguments from evil. But before considering what can be said in support of either of these conclusions, we will need to have some idea of what Frankfurt's objection is and what problems it is supposed to have. So I begin, in section 1, with an account of Frankfurt's counterexample to PAP and of two difficulties it must handle.

1. Two Problems with Frankfurt's Counterexample to PAP

1.1 Frankfurt's Counterexample to PAP
Here is Frankfurt's description of his counterexample:

> Suppose someone – Black, let us say – wants Jones to perform a certain action. Black is prepared to go to considerable lengths to get his way, but he prefers to avoid showing his hand unnecessarily. So he waits until Jones is about to make up his mind what to do, and he does nothing unless it is clear to him (Black is an excellent judge of such things) that Jones is going to decide to do something *other* than what he wants him to do. If it does become clear that Jones is going to decide to do something else, Black takes effective steps to ensure that Jones decides to do, and that he does do, what he wants him to do. Whatever Jones's initial preferences and inclinations, then, Black will have his way. ... Now suppose that Black never has to show his hand because Jones, for reasons of his own, decides to perform and does perform the very action Black wants him to perform.[5]

The key ingredients in the story are an agent who performs the act in question without interference and the presence of another agent, the intervener, who is prepared to intervene to get his way but who doesn't intervene because he doesn't need to in this particular situation. The point of the example is to show that PAP is false. For Jones, in the example, is responsible for performing an action even though he can't do otherwise.

1.2 The Causation Problem
The first problem with Frankfurt's story, from the perspective of at least some incompatibilist agency theorists, is that it tells us that the intervener (Black) would, under certain circumstances, cause Jones to perform some act. The suggestion in the story is that, whether Jones or Black is the cause, it is correct to say that Jones performs the act in question. But incompatibilist agency theorists (at least some of them) will object to this description

of the case in the following way: "It is correct to say that *Jones performs the act* only if Jones is the agent cause of it; if Black causes the act and Jones doesn't agent-cause it, then, strictly speaking, it isn't *Jones* who is performing the act nor is it an act *of Jones's*. Furthermore, due to the nature of agent causation, it is impossible to cause Jones to agent-cause something. So Frankfurt's description of the case is incoherent insofar as it seems to assume that it is possible for Black to cause Jones to perform an act."

The way I'll be thinking of agent causation in this paper is as follows:

> AC. X is the agent cause of e iff each of the following three conditions is satisfied:
> 1. X is a substance that had the power to bring about e
> 2. X exerted its power to bring about e
> 3. nothing distinct from X (not even X's character) caused X to exert its power to bring about e.[6]

Given AC, we can see why it is that the requirement that X be the agent cause of e in order for e to be an act of X's leads to the conclusion that no intervener can cause X to cause e if e is an act of X's.[7]

1.3 The Prediction Problem

The other problem with Frankfurt's story is its claim that Black is able to predict what Jones will do if Black doesn't intervene. The idea seems to be that Black watches Jones carefully, looking for a sign in Jones's behavior prior to t that will make it clear whether Jones will do what Black wants him to do at t (e.g., attempt to kill Smith). Now suppose there is such a sign and it signifies that Jones will try to kill Smith (if there is no intervention). Then presumably the sign would, absent intervention, causally determine Jones's attempting to kill Smith. For if it didn't, then Black has no guarantee that, after the sign and Black's decision not to intervene, Jones won't change his mind and leave Smith alone, much to Black's dismay. But if the sign prior to t *does* causally determine Jones's attempting at t to kill Smith, then that event (Jones's attempting at t to kill Smith) is not an act for which Jones is responsible. In order for it to be an act for which Jones is responsible, Jones must agent-cause it. But if at t Jones agent-causes that event and yet that event is causally determined by the sign Black noticed prior to t, then (assuming this isn't a case of causal overdetermination) Jones's agent-causing at t of that event must be caused by the sign Black noticed prior to t. And that cannot be. For, as we've already noted, according to AC, Jones can't be caused to agent-cause.

In short, either Jones's attempt to kill Smith is causally determined by the sign or not. If not, then the sign gives Black no guarantee that Jones will kill Smith. If it is causally determined, then the event of Jones's attempting to kill Smith isn't an act for which Jones is responsible.[8]

Obviously, the causation problem and the prediction problem are closely related insofar as both are connected with the fact that it is impossible to cause someone to agent-cause. The main difference is that the causation problem has to do with what I'll call 'the counterfactual situation' – the one that would occur if, contrary to fact, Jones were to refrain from the act –

whereas the prediction problem has to do with the actual situation in which Jones does perform the act.

1.4 Solving the Causation Problem

The causation problem can be solved rather easily. What Frankfurt needs is a case where the agent can't *do* otherwise. To guarantee this, we don't need to say that the agent in the counterfactual situation is forced to perform the act that, in the actual situation, he willingly performs (i.e., we don't need to say that the agent is caused to agent-cause what he agent-causes in the actual situation). All we need to guarantee is that the agent can't agent-cause anything other than what he agent-causes in the actual situation. And we can guarantee this by saying that Black has the power temporarily to take away Jones's powers with respect to the act in question. For in that case, Jones can neither perform the act nor intentionally refrain from performing it – indeed, Jones will be temporarily unable to exercise any powers at all with respect to the act in question. Let's say that the act in question is the causing of a volition, V1, to pull the trigger of the gun in his hand. And let's say that Black's plan is this. If Black sees that Jones is about to cause V1 at t, then Black will not intervene. However, if Black sees that Jones is about to intentionally refrain at t from causing V1 (i.e., to cause at t his not causing V1), then Black takes away Jones's powers with respect to V1 at t (so that Jones can neither cause that volition nor cause his not causing it) and Black himself causes V1 at t.[9] In the counterfactual situation, Black doesn't cause Jones to agent-cause V1. Instead, Jones's powers to agent-cause with respect to V1 have been temporarily taken away at t and Black himself causes V1.[10]

In the case just described, Jones is unable to *do* otherwise than cause V1 at t since in the counterfactual situation in which Black takes away Jones's powers and causes V1 himself, Jones isn't able to *do anything* with respect to V1. But despite this fact, the causation problem doesn't arise in connection with this counterfactual situation because there is no suggestion that, in that situation, Jones is caused to agent-cause.[11]

However, solving the causation problem doesn't solve the prediction problem. We've removed the incoherent *causing of agent causation* from the counterfactual situation but we haven't removed it from the actual situation. In section 2, I present a Molinist solution to the prediction problem and exploit it in giving a Molinist Frankfurt-style counterexample to PAP.

2. A Molinist Frankfurt-Style Counterexample

2.1 Molinism and the Prediction Problem

I don't plan on discussing the views of Luis de Molina. The Molinist feature of the Frankfurt-style counterexample I'll be proposing is just that it will involve something very much like middle knowledge, the possibility of which Molina famously endorses. Middle knowledge is knowledge of subjunctive conditionals of freedom. Subjunctive conditionals of freedom are conditionals of the form *If agent X were in circumstances K, X would freely do Y*. It is controversial whether middle knowledge is possible because it is controversial whether there can be true subjunctive conditionals of free-

dom. Although I acknowledge the controversial nature of the assumption that there are true subjunctive conditionals of freedom, I won't make any attempt to defend that assumption here.[12] Instead, I will simply argue for the conditional conclusion that *if there are true subjunctive conditionals of freedom*, then there is a Molinist Frankfurt-style counterexample to PAP *even if* the notion of moral responsibility in PAP is the robust one favored by the incompatibilist agency theorist. Given that it isn't uncommon for incompatibilist agency theorists to agree that there are true subjunctive conditionals of freedom, this conditional conclusion will not be uninteresting.

The attentive reader will notice that I've said my Molinist Frankfurt-style counterexample will involve not middle knowledge but something *very much like* middle knowledge. The difference between the knowledge I have in mind and middle knowledge proper is that the knowledge I have in mind is of subjunctive conditionals of agent causation – i.e., subjunctive conditionals of the form *If agent X were in circumstances K, X would agent-cause Y* where agent causation is understood in accord with AC. Given how common it is for incompatibilist agency theorists to think agent causation is necessary for freedom, it will be natural for them to think the possibility of middle knowledge entails the possibility of the sort of knowledge I have in mind.

The only reason, then, that I call my Frankfurt-style counterexample 'Molinist' is that it will involve knowledge of subjunctive conditionals of agent causation. Nothing more is intended by that name. The suggestion that such knowledge is possible is, of course, as controversial as the suggestion that middle knowledge is possible. One reason that doubts about the possibility of middle knowledge arise is that to some philosophers it doesn't seem possible for subjunctive conditionals of freedom to be true unless something other than the agent causes her free actions (in which case they aren't free actions). Likewise, one reason that doubts about the possibility of knowledge of subjunctive conditionals of agent causation will arise is that to many it doesn't seem possible for subjunctive conditionals of agent causation to be true unless something other than the agent causes her agent-causings (in which case they aren't agent-causings). But middle knowledge sympathizers think that a subjunctive conditional of freedom can be true of an agent even if nothing distinct from that agent causes her free actions. A similar position will be taken by those who think there are true subjunctive conditionals of agent causation (this will include, I assume, all agency theorists who think middle knowledge is possible). They will think that a subjunctive conditional of agent causation can be true of an agent even if nothing distinct from that agent causes her agent-causings. As I said above, I don't want to get into a discussion here of the merits of the view that middle knowledge is possible. I simply want to point out that the view that there are true subjunctive conditionals of agent causation seems to be at least as plausible as the view that there are true subjunctive conditionals of freedom. The reason I point this out is to emphasize that the Molinist Frankfurt-style counterexample I give below (which depends on the possibility of knowledge of subjunctive conditionals of *agent causation*) will be successful if middle knowledge is possible.

It is, no doubt, obvious by now how Molinism can solve the prediction problem. The prediction problem arose because the intervener was sup-

posed to know in advance what Jones would do if there were no interven-
tion. In knowing prior to t what Jones would do at t if there were no inter-
vention, it seemed the intervener had to have knowledge that some event
prior to t would *cause* Jones to perform a certain act at t if there were no
intervention. But since the act in question is supposed to be agent-caused,
it is impossible for the intervener to know *in this way* what Jones would do
at t. However, if the intervener can have prior knowledge of the subjunc-
tive conditionals of agent-causation true of Jones, then the ability to predict
what Jones would do at t if there were no intervention is possible after all.

2.2 The Counterexample Described

Let's say that t^* is a time shortly before t and that the interval of time *from
t^* up until t* includes t^* and all moments after it up until but not including t.
And let's say that circumstances K include a subset of the features of
Jones's environment, but not his own behavior. Now suppose the follow-
ing subjunctive conditional of agent causation is true of Jones:

> A. If from t^* up until t Jones were in circumstances K and Demon
> didn't take away Jones's powers at t with respect to V1, then Jones
> would agent-cause V1 at t.

And suppose that the intervener isn't Black but Demon, a powerful being
with knowledge of subjunctive conditionals of agent causation such as A.
Furthermore, suppose Demon knows (long before t) that Jones will be in
circumstances K from t^* up until t (since Demon plans to arrange for this)
and Demon is considering whether to take away Jones's powers with
respect to V1 at t. Demon wants V1 to occur at t (let's say V1 is the volition
to pull the trigger of the gun in Jones's hand) and would prefer that Jones
agent-cause V1 at t. But, if he knows that Jones won't agent-cause V1 at t,
he will intervene as follows: he will take away Jones's powers at t with
respect to V1 and will cause V1 himself at t.[13] Thus, the following subjunc-
tive conditionals are true of Demon:

> B. If A were true, then Demon would know it (long before t) and
> would refrain from taking away Jones's powers at t with respect to V1.

> C. If A were false, Demon would know it (long before t) and would
> take away Jones's powers at t with respect to V1.

Finally, suppose that, in addition to A-C, it is also true that:

> D. From t^* up until t Jones is in circumstances K and Demon doesn't
> take away Jones's powers at t with respect to V1.

What can we conclude?
   First, we should note that this is *not* a case in which Demon observes some
behavior of Jones's prior to t (such as an inclination to cause V1 at t) which
functions as a sign for Demon that Jones will agent-cause V1 at t. The inter-
vener's prediction that Jones will agent-cause V1 at t doesn't involve his

noticing some behavior of Jones's prior to t that *causally determines* Jones's agent-causing V1 at t. Nor does the intervener's prediction involve knowledge of some behavior of Jones's prior to t which counterfactually implies *without causing* Jones's agent-causing V1 at t.[14] Instead, the intervener's prediction involves knowledge (long before t) of (a) a true subjunctive conditional of agent causation and (b) the truth of the antecedent of that conditional, where that antecedent doesn't describe some behavior of Jones's prior to t but rather the circumstances Jones will be in from t* up until and including[15] t. Thus, in my example, the intervener (Demon) doesn't need to watch Jones carefully as the moment of decision, t, approaches in order to check for some sign indicating what Jones would agent-cause at t (absent intervention). Demon had what he needed long ago, namely, knowledge of (a) and (b).

2.3 Why Jones is Responsible for V1

I now want to argue that the case I've just described in which A-D are true constitutes a counterexample to PAP. It will help if we first consider a slightly different case in which A and D are true but there is no intervener like Demon (so B and C aren't true). In such a case it seems clear that Jones is morally responsible for V1. Jones agent-causes V1 at t and nothing causes him to do so.[16] Clearly the causal buck for V1 stops with Jones. Who, then, besides Jones could be morally responsible for it? Now suppose we alter this case merely by adding to it the existence of Demon of whom B and C are true. Demon has plans to intervene under certain conditions. But those conditions don't obtain and Demon doesn't intervene. For in this Demon case, Jones does exactly what he does in the Demon-less case and he does so with absolutely no interference or influence from Demon. It seems that if Jones is morally responsible for V1 in the Demon-less case, Jones is equally morally responsible for V1 in the Demon case. After all, there has been no intervention. In both the Demon case and the Demon-less case, the causal buck for V1 stops with Jones. Who but he could be morally responsible for V1 in such circumstances? Thus, our first conclusion about this Demon case is that Jones is morally responsible for V1 since in it, just as in the Demon-less case, the causal buck for V1 stops with Jones.[17]

2.4 Why Jones Couldn't Do Otherwise Than Cause V1 at t

Our second conclusion about this case in which A-D are true is that Jones couldn't do otherwise than cause V1 at t. The argument for this conclusion has two parts.[18] In part one, I argue, using only necessary truths as premises, that if

> C. If A were false, Demon would know it (long before t) and would take away Jones's powers at t with respect to V1

is true, then it follows that if

> E. From t* up until t Jones is in circumstances K

is true, it is not the case that

F. At t Jones exercises his power to do otherwise than cause V1 at t.

Given that the conditional *if E then ~F* is entailed by C together with some necessary truths, we may conclude that it is a necessary truth that if C then *if E then ~F*. The second part of the argument uses that conclusion to show that, in the circumstances described in my counterexample, Jones couldn't do otherwise than cause V1 at t.

Let's turn, then, to part one of the argument in which I argue that, necessarily, if C then *if E then ~F*. My strategy here will be to prove first something equivalent to the contrapositive of that conclusion – namely, if both E and F, then ~C – using conditional proof. I will prove that contrapositive by assuming E and F and then using indirect proof to establish ~C (by showing that C leads to a contradiction).

Assume for conditional proof that E and F are true. Since F guarantees the truth of

G. Demon didn't take away Jones's powers at t with respect to V1,

we may conclude that G. We may also conclude that

H. Jones did not agent-cause V1 at t

since F guarantees its truth as well. This gives us the conjunction of E and G and H. But it is a necessary truth that if E and G and H are true then it's not the case that

A. If from t* up until t Jones were in circumstances K and Demon didn't take away Jones's powers at t with respect to V1, then Jones would agent-cause V1 at t.

For A says that if E and G were true then H would be false. So we may conclude (still under the assumption made for conditional proof) that A is false.

Now suppose (for reductio) that

C. If A were false, Demon would know it (long before t) and would take away Jones's powers at t with respect to V1

is true. Without even considering what A says, we can see that it is a necessary truth that if C is true and A is false then G is false (since the consequent of C entails the falsity of G). Since we've proved (using the assumption for conditional proof) that A is false and we are assuming (for reductio) that C, we can use this necessary truth to conclude that G is false. But now we have a contradiction: G is both true (we derived it earlier from F) and false. Since assuming C leads to a contradiction, we may conclude that C is false. And now we may discharge our initial assumption (that E and F are both true) and conclude that if E and F are both true, C is false. From this we may easily derive our desired conclusion: if C is true then *if E then ~F*. And since the only premises used were necessary truths, we may conclude that *necessarily*, if C, then if E then ~F.

Before moving on to consider part two of my argument, it will be help-ful to lay out part one in a more formal manner. I'll begin by laying out the propositions used in the argument:

A. If from t* up until t Jones were in circumstances K and Demon didn't take away Jones's powers at t with respect to V1, then Jones would agent-cause V1 at t.

C. If A were false, Demon would know it (long before t) and would take away Jones's powers at t with respect to V1.

E. From t* up until t Jones is in circumstances K.

F. At t Jones exercises his power to do otherwise than cause V1 at t.

G. Demon didn't take away Jones's powers at t with respect to V1.

H. Jones did not agent-cause V1 at t

Now I can lay out part one of the argument:

1.	E & F	Assume for Conditional Proof
2.	G	From F
3.	H	From F
4.	E & G & H	From 1-3
5.	If (E & G & H), then ~A.	Necessary Truth
6.	~A	From 4 and 5
7.	C	Assume for Indirect Proof
8.	If (~A & C), then ~G.	Necessary Truth
9.	~G	From 6-8
10.	G & ~G	From 2 and 9
11.	~C	7-10 Indirect Proof
12.	If (E & F), then ~C.	1-11 Conditional Proof
13.	If C, then *if E then ~F*.	From 12

Since 13 is derived using only necessary truths as premises, 13 is itself a necessary truth.

Part two of the argument (for the conclusion that, in the counterexample I've described, Jones couldn't do otherwise than cause V1) consists of two applications of the following principle of inference:

PI.   If (i) (If P then Q) and no matter which of her abilities X exer-cised at t, it would be the case that (If P then Q)
and (ii) P and no matter which of her abilities X exercised at t, it would be the case that P,
then (iii) Q and no matter which of her abilities X exercised at t, it would be the case that Q.

PI is very much like the principle Peter van Inwagen calls 'principle β' and

the principle John Martin Fischer calls 'the principle of the transfer of pow-erlessness'.[19] But because it is modeled after the principle proposed by McKay and Johnson, it, like the principle they call 'b4', is not susceptible to the counterexamples that have been proposed to van Inwagen's principle β and to the principle of the transfer of powerlessness.[20]

   The conclusion I want to establish is that, in the counterexample I described, Jones couldn't do otherwise than cause V1. Recall that in the counterexample I described, each of A-D is true. And since

> D. From t* up until t Jones is in circumstances K and Demon doesn't take away Jones's powers at t with respect to V1

is a conjunction with E as its first conjunct, my counterexample also includes the truth of E. Since we are interested in determining what would be true if my counterexample were true, I include both E and C as premis-es in the second part of my argument which proceeds as follows:

<u>First Application of PI</u>
14.   No matter which of his abilities Jones exercised at t, it would be the case that 13.
15.   C
16.   No matter which of his abilities Jones exercised at t, it would be the case that C.
17.   ∴ (If E then ~F) and no matter which of his abilities Jones exer-cised at t, it would be the case that (If E then ~F). [from 13-16 by PI]

<u>Second Application of PI</u>
18.   E
19.   No matter which of his abilities Jones exercised at t, it would be the case that E.
20.   ∴ No matter which of his abilities Jones exercised at t, it would be the case that ~F. [from 17-19 by PI]

As I noted above, premises 15 and 18 are entailed by my description of the Demon case. Premise 14 is uncontroversial since everyone agrees that nothing Jones can do might result in the falsity of a necessary truth. Thus, assuming I'm right in saying that premises 5 and 8 are necessary truths, this leaves only premises 16 and 19 to be defended.

   Premise 16 says that no matter which of his abilities Jones exercised at t, it would be the case that:

> C. If A were false, Demon would know it (long before t) and would at t temporarily take away Jones's powers at t with respect to V1.

In defense of this premise, all I need to do is make explicit an element of the story that is quite naturally taken for granted, namely, that because of Demon's superior powers and the firmness of Demon's plans, Jones isn't able to bring it about that C is false of Demon. Jones simply lacks that sort

of power over Demon's dispositions. Demon is committed to acting in accord with C and there is nothing Jones can do about that.

Premise 19 says that no matter which of his abilities Jones exercised at t, it would be the case that:

E. From t* up until t Jones is in circumstances K.

One might be inclined to defend this premise using the principle called 'the fixity of the past' according to which no one can bring about the falsity of p if p is made true by a hard fact about the past.[21] But Plantinga has argued against this principle by describing scenarios in which a person *can* bring about the falsity of p where p is made true by a hard fact about the past.[22] The key ingredient in his scenarios is a powerful and omniscient being who has foreknowledge which it uses as follows. It foreknows that an agent X will freely do Y at t and, on the basis of that foreknowledge, it permits some event E at t-n, a time long before X was born. But if it had foreknown instead that *X will not freely do Y at t*, it would, on the basis of that foreknowledge, have prevented E from occurring at t-n. Since X's doing Y at t is *free*, we may assume that she could at t bring it about that she freely refrains from Y.[23] But then, given the powerful being's plans and foreknowledge, X could bring it about that E didn't occur at t-n, despite the fact that E's occurring at t-n is a hard fact about the past.

Suppose Plantinga is right about this. Does that prevent me from defending premise 19? No. For I can defend it without relying on the fixity of the past principle to which Plantinga proposes the above counterexample. All I need to do is stipulate that in my Molinist Frankfurt-style counterexample to PAP (i.e., the Demon case) there is no powerful and omniscient being whose permission of the truth of E is dependent on its foreknowledge of something Jones freely does at t. Then I can rely on a weakened version of the fixity of the past principle which says that *absent conditions like those Plantinga describes in his counterexample to the original fixity of the past principle* no one can bring about the falsity of p if p is made true by a hard fact about the past.

The above defense of premises 16 and 19 completes my two-part argument for the second conclusion about the case in which A-D are true, namely, that in that case, Jones doesn't have the ability to do otherwise than cause V1 at t.

2.5 Summary

Now we can combine our two conclusions about the Demon case. The first conclusion, from 2.3, is that Jones is morally responsible for V1 since Jones is responsible in the Demon-less case and the Demon case is relevantly similar. And our second conclusion, from 2.4, is that Jones can't do otherwise than cause V1 at t. Putting these two conclusions together, we have a counterexample to PAP: Jones is morally responsible for causing V1 at t even though Jones couldn't do otherwise than cause V1 at t. Furthermore, this counterexample involves a robust notion of moral responsibility because the responsibility derives from the fact that Jones is the agent cause of V1 and, assuming there could be true subjunctive conditionals of

agent causation, nothing in the description of the counterexample conflicts with the claim that Jones agent-causes V1.[24]

It's worth highlighting the fact that in proposing a counterexample to PAP, I haven't claimed that ability to do otherwise isn't necessary for freedom. My intuitions are fairly strong in support of agent causation, as defined in AC, being sufficient for moral responsibility. That's one main reason why I think the example I've proposed succeeds (assuming there are true subjunctive conditionals of agent causation). But I don't know what to say about freedom. When responsibility and ability to do otherwise come apart, does freedom go with responsibility or with ability to do otherwise? I don't have a firm opinion in answer to this question. (Perhaps there are two incompatibilist notions of freedom: one that goes with responsibility and one that goes with ability to do otherwise.) But since I think the claim that *ability to do otherwise isn't necessary for freedom* is at least questionable, I'll assume, for the purposes of this paper, that ability to do otherwise *is* necessary for freedom.[25]

3. A Difficulty for the Free Will Defense

In this final section, I want to spell out an important consequence my counterexample to PAP has for the Free Will Defense. According to the Free Will Defense, it is possible that God needs to permit evil in order to obtain the good of having significantly[26] free creatures. For in order to have significantly free creatures who do what is right, God must have creatures who are able to do evil (assuming that freedom requires the ability to do otherwise). And in order for them to be able to do evil, God can't be like Demon, the intervener in my Molinist Frankfurt-style counterexample to PAP. In particular, God can't have a plan to prevent creaturely acts (by temporarily removing their powers with respect to the act in question) if he can see via his middle knowledge that, absent such intervention, they would freely do evil. If God *did* have such a plan, then (for the reasons noted in section 2) the creatures wouldn't be able to do otherwise and, hence, wouldn't be significantly free. Thus, if all possible creaturely essences are contingently such that their instantiations would do at least one thing wrong no matter what God did (so long as he instantiated those essences with significant freedom), then God must permit evil if he is to have significantly free creatures.[27]

My purpose isn't to object to the above account of why permitting evil might be necessary for having free creatures.[28] Instead, I want to challenge the assumption that permitting evil is required for obtaining the good that makes having free creatures so valuable. Why is it that having free creatures is thought to be so valuable? I think the main reason is the belief that if God's creatures weren't free, they couldn't engage in a loving relationship with God. Why not? Because love of the sort God has for us and wants from us cannot be caused in a person by anything other than that person. God can't force us to love him because love is the sort of thing for which the lover has to be responsible.

But as soon as we put it that way, we can see (in light of the counterexample from section 2) that the good that God is after – i.e., our loving him in

such a way that we are responsible for it – is compatible with our not being able to do otherwise. Perhaps it is true that *freedom* requires the ability to do otherwise. If so, then the good of our being significantly free may require the permission of evil. But the Molinist Frankfurt-style counterexample to PAP shows that the good of our being *responsible* for the good things we do – in particular, our being responsible for loving God – doesn't require the ability to do otherwise. So, if the good that God is after in making us free is our being responsible for loving him, then there is no need to make us free (and, thereby, risk the moral evil we see around us) in order to get that good. All that is required is that we are responsible for loving God. And God can arrange for that so long as there are at least some subjunctive conditionals of agent causation true of us according to which we agent-cause our love of God in certain circumstances.[29] Then God can arrange for our being responsible for loving him by placing us in those circumstances. And he can, at the same time, prevent us from being able to do otherwise. He can do this by being prepared to intervene in the way Demon was prepared to intervene in Jones's affairs. The only difference is that, whereas Demon intervenes when he knows Jones would agent-cause something other than a particular *wicked* act (i.e., attempting to kill Smith), God intervenes when he sees that we would agent-cause something other than an act of *love* toward God.[30]

But perhaps there is some other good that God can obtain by creating free creatures. Perhaps God wants not just the good of our being responsible for loving him but also some other good for which freedom (with the ability to do otherwise it entails) and not merely responsibility is required. That may be so. I can't rule it out. However, the point I want to make is just that the main reason for thinking that freedom is an important good *fails* to provide God with a good reason to permit evil – or at least it fails if PAP is false. For although our being free does make possible our being responsible for loving God, our being responsible and unable to do otherwise does too. This forces proponents of the Free Will Defense to either defend PAP (and, of course, one could defend it against *my* proposed counterexample by arguing against the possibility of middle knowledge) or come up with some other explanation for why having free creatures is such a valuable thing. Either way, the simple explanation for why freedom is an outweighing good that requires the permission of evil (i.e., the fact that it makes possible a loving relationship with God) cannot stand on its own.[31]

Purdue University

NOTES

1.   Harry Frankfurt, "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy* 66 (1969), 829-39.
2.   See for example David Widerker, "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities," *The Philosophical Review* 104 (1995), 247-61 as well as his paper "Frankfurt's Attack on the Principle of Alternative Possibilities: A Further Look" in *Philosophical Perspectives, 14, Action and Freedom*, ed. James Tomberlin (Boston: Blackwell Publishers, 2000), pp. 181-201. I explain this sort of attack on Frankfurt's counterexample in section 1.3 after first laying out Frankfurt's counterexample to PAP in section 1.1.

3.   I explain the notion of agent causation below in section 1.2. I should note here that there are at least two reasons for thinking that causal determinism is incompatible with freedom and with responsibility: (i) because freedom and responsibility each require the ability to do otherwise and the ability to do otherwise is incompatible with causal determinism and (ii) because, in virtue of involving agent causation, freedom and responsibility each entail that the "causal buck" for the action stops with the agent and *that* is incompatible with causal determinism. I will, in effect, be arguing that reason (i) isn't a good reason for thinking that causal determinism is incompatible with *responsibility*. But this doesn't commit me to the compatibility of causal determinism and responsibility since there is still reason (ii). Nor does it commit me to the failure of reason (i) as a reason for thinking causal determinism is incompatible with *freedom*.

4.   For other attempts to show that there are Frankfurt-style counterexamples to PAP that don't assume the falsity of incompatibilism, see David Hunt, "Frankfurt Counterexamples: Some Comments on the Widerker-Fischer Debate," *Faith and Philosophy* 13 (1996), 395-401; Alfred Mele and David Robb, "Rescuing Frankfurt-style Cases," *Philosophical Review* 107 (1998), 97-112; Derk Pereboom, "Alternative Possibilities and Causal Histories," *Philosophical Perspectives*, 14, *Action and Freedom*, pp. 119-37; and Eleonore Stump, "Libertarian Freedom and the principle of Alternative Possibilities," *Faith, Freedom and Rationality*, eds. Daniel Howard-Snyder and Jeff Jordan (Totowa, NJ: Rowman and Littlefield, 1996), pp. 73-88. For responses to most of these see Widerker's "Frankfurt's Attack on the Principle of Alternative Possibilities: A Further Look."

5.   Frankfurt, pp. 835-36.

6.   I.e., nothing distinct from X caused *on its own* X's exerting its power to bring about e. An event or an agent Y can causally contribute to a result without causing *on its own* that result. If a causal contribution by Z is necessary for a result and the causal contribution by Y to that result isn't causally sufficient for the result and the causal contribution by Y to that result isn't causally sufficient for the result, then Y causally contributes to the result without causing *on its own* that result. However, if Y's causal contribution is causally sufficient for the result, then Y causes *on its own* that result.

7.   AC isn't the only way to define agent causation. Another familiar way to define it is the way that, according to William Rowe (see his paper "The Metaphysics of Freedom: Reid's Theory of Agent Causation," *American Catholic Philosophical Quarterly* LXXIV [2000], 427), Thomas Reid defines it:

AC*. X is the agent cause of e iff each of the following three conditions is satisfied:
   1. X is a substance that had the power to bring about e
   2. X exerted its power to bring about e
   3*. X had the power to refrain from bringing about e.

Rowe argues that 3* entails 3 so that if X is an agent cause of e according to AC*, X is also an agent cause of e according to AC (this is what his response to Objection II on p. 430 of "The Metaphysics of Freedom" amounts to).

8.   See Widerker's "Frankfurt's Attack on the Principle of Alternative Possibilities: A Further Look" for further development of this line of argument.

9.   To cause V1 is to cause the obtaining of the state of affairs 'Jones's willing to pull the trigger of the gun in his hand'. So we can agree with Peter van Inwagen – see his *An Essay on Free Will* (Oxford: Oxford University Press, 1983), pp. 167-70 – that event-particulars (such as 'the fall of the Alamo' or 'the death of Caesar') are individuated by their causes. But this does not prevent us from allowing that V1 can be caused either by Jones or by Black since V1 is an event-universal. (If the incompatibilist agency theorist thinks that X's volitions

or willings aren't the sorts of thing that can be caused by something distinct from X, then the act in the example will have to be the causing of some *other* sort of mental episode – one that isn't essentially agent-caused and that typically leads to the pulling of the trigger of the gun in one's hand.)

10. Jones can be an agent even if his powers with respect to V1 are temporarily taken from him at t. So long as Jones has other powers at t, Jones is still an agent just as I am an agent given the powers I have despite the fact that there are powers I lack.

11. The causation problem could also be solved by saying that if the intervener can see that Jones is about to intentionally refrain from the act the intervener wants him to perform at t, the intervener will destroy Jones at t or render him temporarily comatose and, thereby, unable to do anything at t.

12. For a defense of the possibility of middle knowledge see Thomas Flint, *Divine Providence: The Molinist Account* (Ithaca: Cornell University Press, 1998).

13. Demon prefers this intervention strategy to preventing Jones from being in circumstances K from t* up until t.

14. Fischer suggests this type of example in his "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities," p. 121. See David Widerker and Charlotte Katzoff, "Avoidability and Libertarianism: A Response to Fischer," *Faith and Philosophy* 13 (1996), 415-21 for a response to Fischer; and see Daniel Speak's response to them on Fischer's behalf in "Fischer and Avoidability: A Reply to Widerker and Katzoff." *Faith and Philosophy* 16 (1999), 239-47.

15. I add 'and including' because the antecedent in question (i.e., the antecedent of A) says that Demon doesn't remove Jones's powers with respect to V1 at t – something that occurs, presumably, at t (if it occurs).

16. We can even add that, since there is no intervener like Demon, Jones has at t the power to cause his not agent-causing V1.

17. One might follow Peter van Inwagen in resisting this sort of conclusion on the grounds that Jones can't be responsible for the obtaining of an event-universal E at t if it is inevitable that E occurs at t (see his *Essay on Free Will*, pp. 171-78). However, William Rowe convincingly argues — see his "Causing and Being Responsible for what is Inevitable," *American Philosophical Quarterly* 26 (1989), 153-59 — that van Inwagen is failing to be sensitive to a distinction between two kinds of inevitably occurring event-universals:

> (i) an event-universal E the obtaining of which at t is inevitable because if agent X doesn't bring about the obtaining of E at t, something distinct from X will bring it about at t via a causal process that is initiated if and only if agent X won't bring it about at t
>
> and (ii) an event-universal E the obtaining of which at t is inevitable because there is in place long before t a causal process *completely independent of X's actions and refrainings* that guarantees the obtaining of E at t whether or not X gets involved.

Rowe argues effectively that although van Inwagen is correct that inevitable event-universals of the second kind are ones for which X cannot be responsible, it is, nevertheless, the case that X can be responsible for inevitable event-universals of the first kind. And since V1 in my Demon example is an inevitable event-universal of the first kind, it is one for which Jones can be responsible.

18. Since devising this argument, I've discovered an argument by Fischer (in "Libertarianism and Avoidability: A Reply to Widerker," p. 121) for a similar conclusion. For objections to Fischer's argument (which, as far as I can tell, do not apply to mine) see Widerker and Katzoff, pp. 416-17 and William

Hasker, *The Emergent Self* (Ithaca: Cornell University Press, 1999), pp. 92-93.

19. See van Inwagen's *Essay on Free Will*, section 3.10 and Fischer's "Introduction: God and Freedom" in *God, Foreknowledge and Freedom*, ed. John Martin Fischer (Stanford: Stanford University Press, 1989), pp. 6-7.

20. See Thomas McKay and David Johnson, "A Reconsideration of an Argument against Compatibilism," *Philosophical Topics* 24 (1996), 113-22 for counterexamples to principle β and to the principle of the transfer of power-lessness. Peter van Inwagen now seems to have left principle β behind in favor of a principle more like the one McKay and Johnson defend (see his "Free Will Remains a Mystery" in *Philosophical Perspectives, 14, Action and Freedom*, pp. 1-19). For a defense of yet another principle that is a restricted version of principle β (restricted in order to avoid counterexamples proposed by John Martin Fischer and Mark Ravizza), see Eleonore Stump, "The Direct Argument for Incompatibilism," *Philosophy and Phenomenological Research* LXI (2000), 459-66. Fischer and Ravizza respond to her in their "Replies," *Philosophy and Phenomenological Research* LXI (2000), 467-80.

21. It is notoriously difficult to say what makes a fact about the past a *hard* fact about the past. The basic idea is that it is a fact that is, as Fischer says, "genuinely and solely about the past". So, for example, the fact that Pearl Harbor was bombed in 1941is a hard fact about the past whereas the fact that it was bombed prior to my going to the theater tomorrow to watch the movie *Pearl Harbor* is not (the latter fact is, instead, a *soft* fact about the past, one over which I seem to have control). See Fischer, "Introduction: God and Freedom," pp. 5-6 and Alvin Plantinga, "On Ockham's Way Out," *Faith and Philosophy* 3 (1986), 246-48 for discussion of this issue.

22. See Plantinga's "On Ockham's Way Out," pp. 254-57.

23. Although one of my main conclusions is that *responsibility* doesn't require an ability to do otherwise, I am not arguing that *freedom* doesn't require an ability to do otherwise (see the final paragraph of section 2.5). It's also worth keeping in mind that the knowledge employed in this example by Plantinga is foreknowledge, not middle knowledge. In defense of Plantinga's implicit assumption that foreknowledge is compatible with freedom, see Ted Warfield, "Divine Foreknowledge and Human Freedom are Compatible," *Noûs* 31 (1997), 80-86 and "On Freedom and Foreknowledge: A Reply to Two Critics." *Faith and Philosophy* 17 (2000), 255-59.

24. The reason I focused on subjunctive conditionals of agent causation rather than subjunctive conditionals of freedom is that in my counterexample, Jones isn't able to do otherwise than cause V1. But if the consequent of A were stated in terms of freely causing V1 instead of in terms of agent-causing V1, then, together with D, it would imply that Jones freely caused V1 despite the fact that Jones isn't able to do otherwise than cause V1. This wouldn't sit well with a philosopher who believed that part of what she *means* when she says that a person X is free is that X is able to do otherwise; such a philosopher would say my proposed counterexample is incoherent. But the way I've defined agent-causation (in terms of AC rather than AC* – see note 7) it is clear that it is *not* part of what one *means* when one says that X agent-caused something (in my sense) that X is able to do otherwise. Thus, by focusing on subjunctive conditionals of agent causation (understood in accord with AC rather than AC*) I am able to avoid this sort of charge of incoherence.

25. I lean in the direction of thinking that freedom analytically entails the ability to do otherwise the way being a bachelor entails being an unmarried male. This seems to be the sort of view endorsed by Peter van Inwagen (*Essay on Free Will*, p. 8) and Richard Taylor (see his *Metaphysics*, 4th edition [Englewood Cliffs, NJ: Prentice Hall 1992], pp. 40-44). I've noticed, however,

that those who think of freedom as just *meaning* "whatever is required for being responsible" are put off by the suggestion, in the paragraph to which this note is attached, that it makes sense to think one can be responsible without being free. My only response is to note that my main focus is responsibility and that, when it comes to freedom, I'll be catering to those whose intuitions lend stronger support to "freedom analytically entails the ability to do otherwise" than to "freedom is whatever is required for being responsible".

26. Significant freedom is freedom with respect to acts that are such that performing them is morally right and refraining from them is morally wrong – or vice versa. See Plantinga, *Nature of Necessity* (Oxford: Clarendon Press, 1974), p.166.

27. See Plantinga, *Nature of Necessity*, chapter IX for a detailed development of this point.

28. But see my "Might-Counterfactuals, Transworld Untrustworthiness and Plantinga's Free Will Defence," *Faith and Philosophy* 16 (1999), 336-51.

29. Our depravity would be great indeed if this were not the case. Plantinga's Free Will Defense assumes only the possibility of people being so depraved that no matter what God did, they would do at least *one* thing wrong. But that falls short of the depravity of those who are such that, no matter what God did, they would never do even one loving thing toward God. If all possible creaturely essences were contingently such that their instantiations would be *that* depraved, then its hard to see what reason God would have to instantiate them.

30. Acts of love toward God can be construed rather broadly. It's a familiar Christian teaching that many sorts of acts, in particular acts of kindness toward others, can count as acts of love toward God.

31. Thanks to Jeffrey Brower, William Hasker, Trenton Merricks, Michael Rea, William Rowe and two anonymous referees for helpful comments on earlier drafts. Thanks also to my students, Jennifer Jensen and James Winans, and my colleagues, Jan Cover and Patrick Kain, for helpful discussions concerning the arguments in this paper.