

# Faith and Philosophy: Journal of the Society of Christian Philosophers

---

Volume 16 | Issue 3

Article 3

---

7-1-1999

## Might-Counterfactuals, Transworld Untrustworthiness and Plantinga's Free Will Defence

Michael Bergmann

Follow this and additional works at: <https://place.asburyseminary.edu/faithandphilosophy>

---

### Recommended Citation

Bergmann, Michael (1999) "Might-Counterfactuals, Transworld Untrustworthiness and Plantinga's Free Will Defence," *Faith and Philosophy: Journal of the Society of Christian Philosophers*: Vol. 16 : Iss. 3 , Article 3.

Available at: <https://place.asburyseminary.edu/faithandphilosophy/vol16/iss3/3>

This Article is brought to you for free and open access by the Journals at ePLACE: preserving, learning, and creative exchange. It has been accepted for inclusion in Faith and Philosophy: Journal of the Society of Christian Philosophers by an authorized editor of ePLACE: preserving, learning, and creative exchange.

# MIGHT-COUNTERFACTUALS, TRANSWORLD UNTRUSTWORTHINESS AND PLANTINGA'S FREE WILL DEFENCE

Michael Bergmann

Plantinga's Free Will Defense (FWD) employs the following proposition as a premise:

◇TD. Possibly, every essence is transworld deprived.

I argue that he fails to establish his intended conclusion because the denial of ◇TD is epistemically possible. I then consider an improved version of the FWD which relies on

◇TU. Possibly, every essence is transworld untrustworthy.

(The notion of transworld untrustworthiness is the might-counterfactual counterpart to Plantinga's would-counterfactual notion of transworld depravity.) I argue that the denial of ◇TU is also epistemically possible and, therefore, that the improved FWD fares no better than the original at establishing the compatibility of God and evil.

In his celebrated Free Will Defense (FWD), Alvin Plantinga aims to establish the compatibility of evil and God's existence. His FWD employs the following proposition as a premise:

◇TD. Possibly, every essence is transworld deprived.<sup>1</sup>

I will argue that he fails to establish his intended conclusion because it is reasonable to refrain from believing ◇TD. In other words, I will try to show that it is epistemically possible that<sup>2</sup>

□~TD. Necessarily, some essence or other is *not* transworld deprived.

Notice that one could show that □~TD is epistemically possible either by arguing that it is epistemically possible that

EP1. Some essence is necessarily not transworld deprived

or by arguing that it is epistemically possible that



EP2. No essence is necessarily not transworld depraved but, necessarily, some essence or other is not transworld depraved.

I take the EP1 route.<sup>3</sup>

My paper has four sections. In the first, I defend the epistemic possibility of EP1. In section 2 I explain how Plantinga could sidestep this objection by offering a strengthened version of his FWD - a version that relies on a premise weaker than  $\diamond$ TD. (The weaker premise replaces Plantinga's would-counterfactual notion of transworld depravity with its might-counterfactual counterpart - the notion of transworld untrustworthiness.) Then, in section 3, I respond to this strengthened version of the FWD by defending a strengthened version of EP1. In the final section I summarize the results of the paper in the context of a clarificatory framework that I hope will go some way toward preventing future misunderstanding of Plantinga's oft misunderstood FWD. (Two examples of such misunderstanding, which I discuss briefly in the last section, can be found in John Hick's *Evil and the God of Love* (revised edition)<sup>4</sup> and in J.L. Mackie's *The Miracle of Theism*<sup>5</sup>.)

Before continuing, I should acknowledge that the truth of the claim that EP1 is epistemically possible would not entail that Plantinga's FWD is of no use in responding to the problem of evil. What it would demonstrate is that Plantinga has not shown that the existence of God is compatible with the existence of evil. At most, Plantinga has shown that it is *epistemically possible* that God and evil are compatible (assuming that  $\diamond$ TD, like  $\square\sim$ TD, is epistemically possible). However, even if Plantinga's FWD has shown only that it is epistemically possible that God and evil are compatible, it still manages, thereby, to provide a somewhat satisfactory response to the atheistic charge that the existence of God and the existence of evil are logically incompatible. Such a response is somewhat satisfactory because it shows that the atheistic charge depends on an assumption we can reasonably refrain from accepting. Nevertheless, it is clearly not as satisfactory a response as Plantinga had hoped for.

### 1. In Defense of the Epistemic Possibility of EP1

I will begin by defining the concept of transworld depravity using the notions of an E-perfect world and strong actualization. An *E-perfect world* is a world in which a creaturely essence E's instantiation is significantly free (i.e., has incompatibilist freedom with respect to doing or refraining from doing what is wrong) and always does what is morally right. And God *strongly actualizes* a state of affairs S iff God alone causes S to be actual (thus, S cannot include any creaturely exercise of incompatibilist freedom).<sup>6</sup> With these concepts clear in our minds we can define transworld depravity as follows:

A creaturely essence E is *transworld depraved* (in W) iff for every E-perfect world  $W^*$  the following counterfactual of freedom is true (in W): If God had strongly actualized T of  $W^*$  (the largest state of affairs God strongly actualizes in  $W^*$ ) then E's instantiation would

have freely performed at least one morally wrong act.<sup>7</sup>

Now consider again EP1, the proposition I say is epistemically possible. Can we just see that every essence is such that it is possibly transworld depraved (i.e., can we just see that EP1 is false)? Plantinga says yes (and Rowe, I think, is inclined to agree). I find that surprising. First of all, the concept of transworld depravity is difficult to grasp. But even once it is grasped, I think we need some sort of *reason* for thinking every creaturely essence is possibly transworld depraved. I grant that some necessary truths are such that we can just see that they are true. But this doesn't seem to be one of them. Thus, in the absence of any argument against EP1, it seems only reasonable to say that it is epistemically possible.

Well, perhaps there *is* some reason for thinking that every creaturely essence is possibly transworld depraved. As a matter of fact, I am aware of a line of reasoning that seems to suggest that very conclusion.<sup>8</sup> Here, in two stages, is an argument for the conclusion that every creaturely essence is possibly transworld depraved (i.e., that EP1 is false):

Consider a world *W* in which a creaturely essence *E*1 is significantly free and always does what is right - an *E*1-perfect world. There will be a morally significant action *A* such that *E*1's instantiation freely goes right with respect to *A* in *W*. And since *E*1's instantiation has incompatibilist freedom, there will be some other world *W*\* in which God strongly actualizes *T* of *W* (the largest state of affairs God strongly actualizes in *W*) and *E*1's instantiation goes *wrong* with respect to *A*. But if in *W*\* God *does* strongly actualize *T* of *W* and *E*1's instantiation *does* go wrong with respect to *A*, then, in *W*\*, the following counterfactual of freedom is true: If God had strongly actualized *T* of *W*, *E*1's instantiation *would* have freely gone wrong with respect to *A*. Since the above reasoning applies to any arbitrarily selected essence *E* and any arbitrarily selected *E*-perfect world *W*, we may conclude that

C1. For any creaturely essence *E* and any *E*-perfect world *W* the following counterfactual of freedom is *possibly* true: If God had strongly actualized *T* of *W* (the largest state of affairs God strongly actualizes in *W*), *E*'s instantiation would have freely performed at least one morally wrong act.

Stage two of the argument runs as follows. Consider some essence *E* and let *W*<sub>1-n</sub> be all the *E*-perfect worlds.<sup>9</sup> When we apply the reasoning of stage one to *W*<sub>1</sub>, we reach the conclusion that the following counterfactual of freedom is *possibly* true: If God had strongly actualized *T* of *W*<sub>1</sub>, *E*'s instantiation would have freely performed at least one morally wrong act. Using similar reasoning for all the *E*-perfect worlds, we may conclude that each of the following counterfactuals of freedom is *possibly* true:

CF1. If God had strongly actualized *T* of *W*<sub>1</sub>, *E*'s instantiation would have freely performed at least one morally wrong act.

CF2. If God had strongly actualized T of  $W_2$ , E's instantiation would have freely performed at least one morally wrong act.  
 CF<sub>n</sub>. If God had strongly actualized T of  $W_n$ , E's instantiation would have freely performed at least one morally wrong act.

Now there are some counterfactuals of freedom, each of which is possibly true, that could not be true in the same world. For example, we may suppose that each of the following counterfactuals of freedom is possibly true:

If situation S were actual, E's instantiation would freely do A.  
 If situation S were actual, E's instantiation would freely do not-A.

Nevertheless, they cannot both be true in the same world. Now what about CF1 through CF<sub>n</sub>? We have concluded that each is possibly true. But can all of them be true in the same world? It seems so. For no two of them will be of the form just noted. That is, no two will specify the same situation in the antecedent and also specify contradictory actions in the consequent.

Thus, we may conclude that it is possible that all of CF1 through CF<sub>n</sub> are true, i.e., there is a single world in which all of them are true. But then there is some world  $W^*$  in which for every E-perfect world  $W$  the following counterfactual of freedom is true: If God had strongly actualized T of  $W$  (the largest state of affairs God strongly actualizes in  $W$ ), E's instantiation would have freely performed at least one morally wrong act. In other words, it is possible that E is transworld depraved. And since C1 applies to any arbitrarily selected creaturely essence, we may conclude that

C2. Every essence is possibly transworld depraved.<sup>10</sup>

C2 entails that EP1 is false so this argument gives us a reason for thinking that EP1 is not epistemically possible.<sup>11</sup>

In light of the above argument (which I will henceforth call "the argument against EP1"), one might think that we *can* see that every essence is possibly transworld depraved. Thus, if I want to maintain that EP1 is epistemically possible, I will have to show what is wrong with that argument.

We need look no further than stage one. Notice that that stage relied on the following premise:

P1. If in  $W^*$  God *does* strongly actualize T of  $W$  and E1's instantiation *does* freely perform at least one morally wrong act, then, in  $W^*$ , the following counterfactual of freedom is true: If God had strongly actualized T of  $W$ , E1's instantiation *would* have freely performed at least one morally wrong act.

But Plantinga explicitly rejects such a premise (at least as a general principle).<sup>12</sup> He insists that *A & B* does not imply *If A were the case, then B would be the case*. For it may be that although A and B are true, B is not true in certain of the nearby A-worlds. And the counterfactual in question is true in an A-world W only if B is true in W and in *all* the other A-worlds closest to W. (The assumption here is that W might be tied for closeness with itself with other nearby A-worlds.<sup>13</sup>) Thus, the above argument against EP1 will not salvage Plantinga's FWD, at least not from Plantinga's perspective.

## 2. Revising Plantinga's FWD

At this point, there are a variety of responses available to Plantinga. However, when someone objects to a premise in your argument, it is always nice to point out that you can revise your argument so that it doesn't require that premise. Just such a response is available to Plantinga at this juncture. He doesn't need to rely on P1 in order to argue that God and evil are compatible. Instead he can rely on

P1\*. If in W\* God *does* strongly actualize T of W and E1's instantiation *does* freely perform at least one morally wrong act, then, in W\*, the following counterfactual of freedom is true: If God had strongly actualized T of W, E1's instantiation *might* have freely performed at least one morally wrong act.

Notice that P1\* is not open to objection in the way P1 is. For the might-counterfactual *If A were the case B might be the case* obviously does follow from the conjunction of A and B. (I'm reading the might-counterfactual *If A were the case B might be the case* as being true just in case B is true in at least one of the closest A-worlds.<sup>14</sup>) Thus, by modifying the argument so that it relies on P1\* instead of P1, Plantinga would be able to sidestep the objection I proposed at the end of section 1.

Of course if we make this change to the section 1 argument against EP1, we are forced to make a few others as well. The conclusion of stage one will have to be changed from C1 to

C1\*. For any creaturely essence E and any E-perfect world W the following counterfactual of freedom is *possibly* true: If God had strongly actualized T of W (the largest state of affairs God strongly actualizes in W), E1's instantiation *might* have freely performed at least one morally wrong act.

And the conclusion of stage two will have to be changed from C2 to

C2\*. Every essence is possibly transworld *untrustworthy*

where transworld untrustworthiness is defined as follows:

A creaturely essence *E* is transworld *untrustworthy* (in *W*) iff for every *E*-perfect world *W*\* the following counterfactual of freedom is true (in *W*): If God had strongly actualized *T* of *W*\* (the largest state of affairs God strongly actualizes in *W*\*) then *E*'s instantiation *might* have freely performed at least one morally wrong act.

(This definition differs from the definition of transworld depravity only in the parts that are italicized.)

But now we have a problem. Revising the argument against EP1 so that it relies on the more plausible P1\* instead of the questionable P1 makes it cease to be an argument against EP1. For though C2 is incompatible with EP1, C2\* is not. Instead, C2\* is incompatible with

EP1\*. Some essence is necessarily not transworld untrustworthy.

So now we have two different arguments. Let's call this modification of the argument against EP1 (which replaces P1, C1 and C2 with P1\*, C1\* and C2\*) "the argument against EP1\*." I said earlier that Plantinga could argue for his intended conclusion without relying on P1. It's true that the argument against EP1\* avoids relying on P1. But how does it help him defend his intended conclusion? The argument against EP1 helped insofar as it would have shown that EP1 is not epistemically possible (thereby deflecting my objection to  $\diamond$ TD). But what exactly is the relevance of the argument against EP1\*?

The answer is that the argument against EP1\* is helpful to the proponent of the FWD only if she revises the FWD in the way I will now explain. Plantinga's FWD takes for granted a principle something like the following:

P. For any essence *E*, any *E*-perfect world *W* and any action *A* that is morally significant for *E*'s instantiation in *W*, one of the following counterfactuals of freedom is true:

(i) If God had strongly actualized *T* of *W*, *E*'s instantiation *would* have freely gone *wrong* with respect to *A*.

or

(ii) If God had strongly actualized *T* of *W*, *E*'s instantiation *would* have freely gone *right* with respect to *A*.

Now Plantinga says that this is an assumption that can be dispensed with in giving the FWD.<sup>15</sup> He doesn't do so when he employs  $\diamond$ TD as a premise. But, as we will see, he is right that he could dispense with that assumption.

How exactly would the FWD go if it did not take the above principle for granted? Well, instead of starting with  $\diamond$ TD as a premise, the proponent of the strengthened version of the FWD would start with

$\diamond$ TU. Possibly, every essence is transworld untrustworthy.

Notice that if an essence *E* is transworld untrustworthy, God cannot

weakly actualize an E-perfect world. That is, there is nothing God could have done - no series of actions God could have taken - such that if he had, an E-perfect world *would* have been actual (notice the emphasis). Consequently, if *every* essence is transworld untrustworthy, then there is no way God could create free creatures and *ensure* that no evil occurs.

But, you might reply, God could have *superweakly* actualized an E-perfect world where

God superweakly actualizes W iff (i) God strongly actualizes T of W, (ii) W is actual but (iii) it's not the case that if God had strongly actualized T of W, W *would* have been actual.

That is true, but it doesn't help the atheologian who wants to show that God's existence is incompatible with the existence of evil. For God cannot be blamed for failing to superweakly actualize an E-perfect world if E is transworld untrustworthy. It's true that if E is transworld untrustworthy God can strongly actualize a state of affairs such that if he does so, E's instantiation *might* always freely do what is right. But there is nothing God could do that would *secure* the actualization of an E-perfect world. Thus, if every essence is transworld untrustworthy, then no matter which essences God instantiates, he cannot *guarantee* that no free creature will do what is wrong. It follows from  $\diamond$ TU (and the rest of the original FWD), therefore, that God's existence is compatible with the existence of evil. This is an improved version of Plantinga's FWD insofar as it does the same work as the original and yet relies on a weaker premise.

So Plantinga doesn't have to rely on  $\diamond$ TD as a premise in his FWD. He can instead rely on  $\diamond$ TU.<sup>16</sup> To oppose *this* premise I will have to maintain that it is epistemically possible that

$\square\sim$ TU. Necessarily, some essence or other is *not* transworld untrustworthy.

And, as is the case with  $\square\sim$ TD, there are two ways that one could show that  $\square\sim$ TU is epistemically possible. One could show that it is epistemically possible that

EP1\*. Some essence is necessarily not transworld untrustworthy.<sup>17</sup>

Or one could argue that it is epistemically possible that

EP2\*. No essence is necessarily not transworld untrustworthy but, necessarily, some essence or other is not transworld untrustworthy.

Now we can see the relevance of the argument against EP1\*. If it is successful, it prevents someone like myself from relying on the epistemic possibility of EP1\* in an attempt to object to the *revised* FWD. And since the argument against EP1\* manages to avoid relying on the questionable premise P1, the revised FWD emerges unscathed from the objections considered thus far.



3. *In Defense of the Epistemic Possibility of EP1\**

In section 2 I explained how the section 1 objection to stage *one* of the argument against EP1 does not work against stage one of the argument against EP1\*. But is stage *two* of the argument against EP1\* successful? I will argue that the denial of one of the premises of stage two is epistemically possible. If I'm right about this, the argument against EP1\* will be of no use in challenging the claim that EP1\* is epistemically possible. For EP1\*, like EP1, is a modal proposition whose falsity we cannot just see. Unless we have some reason (that doesn't depend on a premise whose falsity is epistemically possible) for thinking EP1\* is false, it is reasonable to take EP1\* to be epistemically possible.

So what's wrong with stage two of the argument against EP1\*? Recall that stage two of the argument against EP1 employed the following premise:

P2. For any essence E, if each of CF1 through CF<sub>n</sub> is possibly true then it is possible that each of CF1 through CF<sub>n</sub> is true.

In the argument against EP1\*, instead of CF1 through CF<sub>n</sub> (which are *would*-counterfactuals of freedom) we have

CF1\*. If God had strongly actualized T of W<sub>1</sub>, E's instantiation *might* have freely performed at least one morally wrong act.

CF2\*. If God had strongly actualized T of W<sub>2</sub>, E's instantiation *might* have freely performed at least one morally wrong act.

...

CF<sub>n</sub>\*. If God had strongly actualized T of W<sub>n</sub>, E's instantiation *might* have freely performed at least one morally wrong act.

(Recall that W<sub>1-n</sub> are *all* the E-perfect worlds.<sup>18</sup>) Thus, in the argument against EP1\* we must replace P2 with

P2\*. For any essence E, if each of CF1\* through CF<sub>n</sub>\* is possibly true then it is possible that each of CF1\* through CF<sub>n</sub>\* is true.

I will be questioning P2\*. But before I do, I would like to point out, in support of it, that although it is easy to see how two *would*-counterfactuals could be incompatible, it is not so easy to see how two *might*-counterfactuals could be incompatible. As I said earlier, the following two *would*-counterfactuals are obviously incompatible:

If situation S were actual, E's instantiation would freely do A.

If situation S were actual, E's instantiation would freely do not-A.

But the same cannot be said for the corresponding *might*-counterfactuals:

If situation S were actual, E's instantiation might freely do A.

If situation S were actual, E's instantiation might freely do not-A.

Only a would-counterfactual is obviously incompatible with a might-counterfactual (e.g., the first of the two might-counterfactuals is obviously incompatible with the second of the two would-counterfactuals).

This helps us to see what's involved in denying P2\*. Consider the following list of counterfactuals of freedom (where the notation used is the same as in the two earlier lists):

CF1\*\*. If God had strongly actualized T of  $W_1$ , E's instantiation *would* have always freely done what is *right*.

CF2\*\*. If God had strongly actualized T of  $W_2$ , E's instantiation *would* have always freely done what is *right*.

...

CFn\*\*. If God has strongly actualized T of  $W_n$ , E's instantiation *would* have always freely done what is *right*.

Using this list, we can state the denial of P2\* as follows:

EP3. For some essence E, each of CF1\* through CFn\* is possibly true but it is necessarily the case that at least one of CF1\*\* through CFn\*\* is true.

In order to challenge P2\*, I will have to say that EP3 is epistemically possible.

The previous two paragraphs may make it seem difficult to question stage two of the argument for EP1\*. For resisting that argument (in the way I suggest) commits us to conceding the epistemic possibility of its being *necessary* that at least one of a list of counterfactuals of freedom whose consequents assert that an essence E's instantiation *would have always gone right* is true (note that this is not to say that there is one counterfactual on such a list that is necessarily true). But what could possibly explain the truth of EP3?

As I said, the above paragraphs may make it *seem* difficult to question stage two of the argument for EP1\*. But I don't see why they *should* make questioning it difficult. It's true that I don't know what could explain the truth of EP3. But of course I don't need to *explain* the truth of EP3 in order to defend the modest claim that it is epistemically possible. I grant that I can't *see* that EP3 is true. But a failure to see that EP3 is true is not the same thing as seeing that it is false.<sup>19</sup> EP3 is like EP1 and EP1\* in that it is a modal proposition whose truth or falsity is not something we can just see in the absence of reasons. So until we have reasons for rejecting EP3, we should concede that it is epistemically possible. And because EP3 is epistemically possible, we are left with the conclusion that, in the absence of any other arguments for the falsity of EP1\*, EP1\* is epistemically possible too. This is enough to show that it is epistemically possible that  $\diamond TU$  is false (and epistemically possible that  $\diamond TD$  is false since  $\diamond TD$  obviously entails  $\diamond TU$ ). Thus, neither Plantinga's FWD nor the improved version of the FWD that I've proposed here is successful in showing that God and evil are compatible. At best they show that it is *epistemically possible* that God and evil are compatible.

The main sort of resistance I've encountered to this conclusion has to do with the fact that the truth of EP3 would be extremely odd. It is extremely odd to think that the counterfactuals of freedom in question are logically dependent on one another in the way EP3 suggests. I grant that that seems odd. But my point is that the oddness is at best a reason to refrain from endorsing EP3. The oddness of EP3 plays a role in our failing to see that it is true. But it does not enable us to see that EP3 is false. Consequently, EP3 is epistemically possible.

#### 4. Clarificatory Summary

In this final section, I provide a framework within which to understand Plantinga's FWD as well as the objections and refinements discussed above. The FWD has often been misunderstood. I hope that what I say here can help prevent further misunderstanding in the future. After laying out this framework, I will use it to explain the concepts of transworld depravity and transworld untrustworthiness. I will also use it to summarize the arguments against EP1 and EP1\* (and for  $\diamond$ TD and  $\diamond$ TU) as well as some objections to both the original and improved versions of Plantinga's FWD.

Let's suppose we can number all those creaturely essences which in some world or other are such that their instantiations have significant freedom and never do wrong. Let's call them E1-En.<sup>20</sup> Now consider E1. Presumably there will be many worlds in which E1's instantiation has significant freedom and never does wrong. These are E1-perfect worlds. Let's number them E1PW<sub>1</sub>-E1PW<sub>n</sub>. Let T of a world be the largest state of affairs God strongly actualizes in that world. Now consider the following lists of counterfactuals of freedom ('WCF' for would-counterfactuals and 'MCF' for might-counterfactuals):

##### List WCF1

If God had strongly actualized T of E1PW<sub>1</sub>, E1's instantiation *would* have freely performed at least one morally wrong act.

If God had strongly actualized T of E1PW<sub>2</sub>, E1's instantiation *would* have freely performed at least one morally wrong act.

...

If God had strongly actualized T of E1PW<sub>n</sub>, E1's instantiation *would* have freely performed at least one morally wrong act.

##### List MCF1

If God had strongly actualized T of E1PW<sub>1</sub>, E1's instantiation *might* have freely performed at least one morally wrong act.

If God had strongly actualized T of E1PW<sub>2</sub>, E1's instantiation *might* have freely performed at least one morally wrong act.

...

If God had strongly actualized T of E1PW<sub>n</sub>, E1's instantiation *might* have freely performed at least one morally wrong act.

Notice that there will be corresponding lists WCF2 and MCF2 and so on

right on up to WCF<sub>n</sub> and MCF<sub>n</sub> (one pair of lists for each of the essences E<sub>1</sub>-E<sub>n</sub> mentioned above). Lists WCF<sub>2</sub> and MCF<sub>2</sub> will refer to the E<sub>2</sub>-perfect worlds E<sub>2</sub>PW<sub>1</sub>-E<sub>2</sub>PW<sub>n</sub> and lists WCF<sub>3</sub> and MCF<sub>3</sub> will refer to E<sub>3</sub>PW<sub>1</sub>-E<sub>3</sub>PW<sub>n</sub>. More generally, lists WCF<sub>i</sub> and MCF<sub>i</sub> will refer to E<sub>i</sub>PW<sub>1</sub>-E<sub>i</sub>PW<sub>n</sub> (defined in a way parallel to the above definitions of E<sub>1</sub>PW<sub>1</sub>-E<sub>1</sub>PW<sub>n</sub>).

The above lists enable us to obtain a better grasp of some of the complicated properties and propositions considered in this paper. A significantly free creaturely essence E<sub>i</sub> is transworld depraved if and only if *each* member of list WCF<sub>i</sub> is *in fact* true. And  $\diamond$ TD (the questionable premise of Plantinga's original FWD) is true if and only if there is *some* world in which *every* member of *every* WCF-list is true. Likewise, a significantly free creaturely essence E<sub>i</sub> is transworld untrustworthy if and only if *each* member of list MCF<sub>i</sub> is *in fact* true. And  $\diamond$ TU (the questionable premise of the improved version of Plantinga's FWD) is true if and only if there is *some* world in which *every* member of *every* MCF-list is true. It's true that in order to *fully* understand the properties and propositions just explained one must put forth some effort to understand the notation as it is used in the WCF- and MCF-lists. But what is immediately obvious from these explanations is that the properties and propositions in question have to do with lists of counterfactuals of freedom. This is a significant gain in elucidation. For it's been my experience that philosophers who discuss these properties and propositions fail to recognize this.

Let me illustrate this failure with a couple of examples of misunderstanding (the first is Hick, the second is Mackie):

For it may be that the beings (or some of them) who freely go wrong in the existing world would also go wrong in any other possible world. They would then exhibit what Plantinga calls 'transworld depravity'.<sup>21</sup>

Curley Smith suffers from what Plantinga calls 'transworld depravity': in whatever world he exists, if he is significantly free he commits some wrong actions: this Plantinga takes to be a fact about Curley's individual essence.<sup>22</sup>

Each takes transworld depravity to be an *essential* property of essences: the property of being such that if it is instantiated with freedom, the resulting creature will freely perform at least one morally wrong act. But transworld depravity is a *contingent* property of essences: it's a property an essence has if a certain list of would-counterfactuals of freedom pertaining to that essence's instantiation is *in fact* true.<sup>23</sup>

Furthermore, the antecedent of each counterfactual of freedom on the list makes reference to a possible world in which the essence in question is *perfect* (recall that the counterfactuals say that if God had actualized T of some E-perfect world, then E's instantiation would have done something wrong). But then it is obvious that an essence suffering from transworld depravity is perfect in some world (and, therefore, obvious that Hick's and Mackie's understanding of transworld depravity is mis-

taken). And the fact that the instantiation of an essence that *in fact* suffers from transworld depravity is *perfect* in some world also makes it evident that transworld depravity is a *contingent* property. For if an essence E's instantiation is perfect in some world W, then in W the following is obviously true: there is some E-perfect world (i.e., W itself) such that if God had strongly actualized T of it, it is *not* the case that E's instantiation would have performed at least one morally wrong act. But that is to say that in W, E is not transworld deprived (since it shows that at least one member of the list of counterfactuals in question is false).

I suspect that one of the main reasons for the misunderstandings mentioned above is the name of the property: *transworld* depravity. The name suggests that it is a property exemplified in all worlds by (and therefore essential to) the essences that have it. But transworld depravity does not entail that the essence that has it is deprived (i.e., performs some morally wrong act) in every world. Rather it entails that the essence that has it is contingently such that it would have been deprived *no matter which world* God had tried to actualize (by strongly actualizing T of that world).

Let's return to laying out the clarificatory framework. Keeping in mind the meaning of the notation used in the WCF- and MCF-lists, consider the following list of *conjunctions* ('F' for *freely* going wrong):

*List F1*

T of E1PW<sub>1</sub> is actual *and* E1's instantiation freely performs at least one morally wrong act.

T of E1PW<sub>2</sub> is actual *and* E1's instantiation freely performs at least one morally wrong act.

...

T of E1PW<sub>n</sub> is actual *and* E1's instantiation freely performs at least one morally wrong act.

Again, notice that there will be corresponding lists F2, F3 and so on right up to F<sub>n</sub> for each of the essences E1-E<sub>n</sub> mentioned earlier.

With these lists in mind, we can summarize the arguments and objections discussed in sections 1-3. Both the argument against EP1 and the argument against EP1\* began with the stipulation that the essences in question have incompatibilist freedom and pointed out that that stipulation entails

(I) Each member of each F-list is true in some world or other.

All parties involved in this discussion grant this entailment (for the purposes of discussion if not because they think it is true). The argument for  $\diamond$ TD via the denial of EP1 runs as follows:

Each of the essences we are considering has incompatibilist freedom. Therefore,

(I) Each member of each F-list is true in some world or other.

Therefore,

- (II) Each member of each WCF-list is true in some world or other.
- Therefore,
- (III) Each WCF-list is such that there is some world in which *all* its members are true (i.e., EP1 is false).
- Therefore,
- (IV) There is some world in which all members of all WCF-lists are true (i.e.,  $\diamond$ TD).

In a recent paper, Howard-Snyder and O'Leary-Hawthorne object to the inference from (III) to (IV) (and in a reply to their paper, William Rowe defends this inference).<sup>24</sup> But they seemed to grant P2, which is the inference from (II) to (III). In section 1 I objected to P1 which is the inference from (I) to (II).<sup>25</sup>

But then I pointed out that the improved version of Plantinga's FWD relies on  $\diamond$ TU not  $\diamond$ TD. The argument for  $\diamond$ TU via the denial of EP1\* goes like this:

- Each of the essences we are considering has incompatibilist freedom. Therefore,
- (I) Each member of each F-list is true in some world or other.
- Therefore,
- (V) Each member of each MCF-list is true in some world or other.
- Therefore,
- (VI) Each MCF-list is such that there is some world in which all its members are true (i.e., EP1\* is false).
- Therefore,
- (VII) There is some world in which all members of all MCF-lists are true (i.e.,  $\diamond$ TU).

I granted P1\*, the inference from (I) to (V), and objected to P2\*, the inference from (V) to (VI). (Neither I nor Howard-Snyder and O'Leary-Hawthorne had anything to say about the inference from (VI) to (VII) though presumably, Howard-Snyder and O'Leary-Hawthorne would want to resist it.)

That completes my summary of the arguments and objections. However, before closing I will address one final worry. I say Plantinga's FWD is improved by employing the weaker  $\diamond$ TU rather than the stronger  $\diamond$ TD as a premise. And in a way that's true (since the same conclusion is established using fewer assumptions). But there is also an advantage to sticking with would-counterfactuals rather than moving to a discussion of might-counterfactuals.<sup>26</sup> For it is the atheist objector who needs would-counterfactuals; it is by means of God's knowledge of would-counterfactuals of freedom that God is supposed to be able to actualize a world with free creatures who always freely do what is right. Thus, by granting principles like P (see section 2 of this paper), Plantinga is giving the atheistic objector all her assumptions and only *then* responding. And that of course makes good sense dialectically.

But what of it? How does this create a problem for anything I've said in this paper? Well, it suggests that the recommended improvement of Plantinga's FWD isn't advisable from a certain perspective. Worse, it suggests that the only objection I considered to Plantinga's original FWD (that EP1 is epistemically possible because the argument against it relied on P1 which is questionable) would not be one the atheistic objector would be keen to employ. Let me rectify that by giving another objection to the original FWD.<sup>27</sup>

In section 3 I explained why it was reasonable to refrain from accepting the inference from (V) to (VI): basically it's because we can't *see* that this inference is valid (though we can't see it is invalid either). Because it is reasonable to refrain from accepting this inference, the argument against EP1\* is not successful. Hence EP1\* and, therefore,  $\sim\Diamond TU$  are epistemically possible. That was my objection to the revised FWD. But a parallel objection applies to the original FWD. For P2, the inference from (II) to (III), is questionable in the same way as is P2\*, the inference from (V) to (VI). Because it is reasonable to refrain from accepting the inference from (II) to (III), the argument against EP1 doesn't succeed. Consequently, EP1 and, therefore,  $\sim\Diamond TD$  are epistemically possible.<sup>28</sup>

Purdue University

#### NOTES

1. See chapter IX of Plantinga's *The Nature of Necessity* (Oxford: Clarendon Press, 1974).

2. I am using the claim 'It is epistemically possible that p' as follows: It is epistemically possible that p iff it is reasonable to refrain from believing  $\sim p$ . (Thus, certain necessary falsehoods - for example, either Goldbach's conjecture or its denial - are epistemically possible.)

3. In their paper "Transworld Sanctity and Plantinga's Free Will Defense", *International Journal for Philosophy of Religion* 44 (1998), 1-21, Daniel Howard-Snyder and John O'Leary-Hawthorne propose a similar objection to Plantinga's FWD only they *seem* to take the EP2 route. (That they take the EP2 route is suggested by their employment of David Lewis' principle of plenitude near the end of their section 2; see their paper for details.) William Rowe has defended Plantinga's FWD against their objection in his paper "In Defense of 'The Free Will Defense'", *International Journal for Philosophy of Religion* 44 (1998), 115-20. It's worth noting that Rowe's response to their objection does not present a challenge to *my* objection to Plantinga's FWD. For his response explicitly relies on the assumption that Howard-Snyder and O'Leary-Hawthorne will *not* take the EP1 route. See Rowe, note 4.

4. San Francisco: Harper and Row, 1978.

5. Oxford: Clarendon Press, 1982.

6. Plantinga employs the notion of strong actualization in *The Nature of Necessity* (see p. 173).

7. See *The Nature of Necessity*, p. 188. This is not how Plantinga states the definition of transworld depravity but it is basically equivalent to his definition.

8. It was suggested to me by William Rowe though he should not be charged with proposing the specific arguments I criticize in this paper.

9. Here and elsewhere in this paper I will assume that there are finitely

many worlds of this sort when in fact there are probably infinitely many. However, it is a little bit easier to present my arguments on the assumption that the number of worlds in question is finite. And since nothing depends on this simplifying assumption, I won't bother to further complicate an already difficult paper by avoiding it.

10. Notice that C2 is weaker than Plantinga's  $\diamond$ TD. The move from C2 to  $\diamond$ TD is resisted by those taking the EP2 route in objecting to Plantinga's FWD. See section 2 of Howard-Snyder and O'Leary-Hawthorne.

11. Remember, this is not what I think. I am merely presenting an argument to which I will later be objecting.

12. See his "Respondeo" in *Warrant in Contemporary Epistemology: Essays in Honor of Plantinga's Theory of Knowledge*, ed. Jonathan Kvanvig (Lanham, Maryland: Rowman and Littlefield, 1996), pp. 328-29.

13. See David Lewis, *Counterfactuals* (Oxford: Basil Blackwell, 1973), p. 29 for a discussion of this assumption. It should be noted that Lewis himself accepts the inference from *A & B to If A were the case then B would be the case*.

14. See Lewis, pp. 21 & 29. It's important to recognize that I'm not reading the consequents of might-counterfactuals as assertions of epistemic possibility. Thus, the counterfactual *If A were the case B might be the case* is not to be understood as saying *If A were the case it is epistemically possible that B would be the case*.

15. *The Nature of Necessity*, p. 180.

16. If he did, Howard-Snyder's and O'Leary-Hawthorne's objection to Plantinga's FWD would have to be modified. Their objection to  $\diamond$ TD relies on the incompatibility of transworld depravity and transworld sanctity, the latter of which they define as follows:

An essence E is transworld *sanctified* (in W) iff for every E-perfect world  $W^*$  the following counterfactual of freedom is *false* (in W): If God had strongly actualized T of  $W^*$  (the largest state of affairs God strongly actualizes in  $W^*$ ) then E's instantiation would have freely performed at least one morally wrong act.

(They don't use these exact words; see section 2 of Howard-Snyder and O'Leary-Hawthorne.) But transworld sanctity is not incompatible with transworld *untrustworthiness* - or at least we have no good reason to think so. They certainly can't *prove* it is in the way they prove the incompatibility of transworld sanctity and transworld depravity. For it's hard to see how the truth of some might-counterfactual (of the sort that will be true of an essence that is transworld untrustworthy) entails the truth of any would-counterfactual (of the sort that will be false of an essence that is transworld sanctified). In order to give their style of objection to  $\diamond$ TU, they would have to speak instead of transworld *supersanctity* where

An essence E is transworld *supersanctified* (in W) iff for every E-perfect world  $W^*$  the following counterfactual of freedom is *true* (in W): If God had strongly actualized T of  $W^*$  (the largest state of affairs God strongly actualizes in  $W^*$ ) then E's instantiation would have *always freely done what is right*.

For they *can* prove that transworld supersanctity is incompatible with transworld untrustworthiness. Or perhaps they are taking for granted the truth of principle P (see my section 2) which entails that transworld sanctity is logically equivalent to transworld supersanctity.

17. Notice that you can't show that EP1\* is epistemically possible by relying on the incompatibility of transworld untrustworthiness and transworld super-



sanctity (see note 16 for a definition of the latter). For though they *are* incompatible, we can prove that no essence is necessarily transworld supersanctified. To see this note that the claim that an essence *is* necessarily transworld supersanctified conflicts with C1\*, the conclusion of stage one of the argument against EP1\*. And that stage of the argument against EP1\* is sound. (So, in responding to the *revised* version of Plantinga's FWD, Howard-Snyder and O'Leary-Hawthorne are *forced* to take the EP2\* route whether they want to or not. For their sort of response requires them to rely on the incompatibility of transworld supersanctity and transworld untrustworthiness.)

Rowe mentions that William Hasker has pointed out to him that "it is relatively easy to show that each creaturely essence is such that in some world it is not transworld sanctified" (see Rowe, note 4). I'm doubtful of that. For it is sufficient for an essence's being necessarily transworld sanctified (see my note 16 for a definition of that property) that *no* would-counterfactuals of freedom are true of it in any world. Thus, to show what Hasker claims is *easy* to show requires showing that there are true would-counterfactuals of freedom - a controversial claim at best. But, as I explained in the previous paragraph, it *is* relatively easy to show that each creaturely essence is such that in some world it is not transworld *supersanctified*. (Perhaps Hasker too was taking for granted principle P which entails the logical equivalence of transworld sanctity and transworld supersanctity.)

18. See note 9 in which I address the worry that there are infinitely many E-perfect worlds, not finitely many as I assume in the text.

19. This is something that seems to be ignored in the section 1 argument against EP1 when it says

Now what about CF1 through CFn? We have concluded that each is possibly true. But can all of them be true in the same world? It seems so. For no two of them will be of the form just noted. That is, no two will specify the same situation in the antecedent and also specify contradictory actions in the consequent.

The (erroneous) point being made here is that because we *can't* see that these counterfactuals are *incompatible* we *can* see that they are *compatible*.

20. The comments made in note 9 concerning the simplifying assumption that there are finitely many worlds of a certain sort also apply here to the assumption that there are finitely many essences of a certain sort.

21. John Hick, *Evil and the God of Love (revised edition)* (San Francisco: Harper and Row, 1978), p. 367.

22. J.L. Mackie, *The Miracle of Theism* (Oxford: Clarendon Press, 1982), p. 174.

23. A more recent example of a philosopher misunderstanding Plantinga's FWD is Douglas Geivett's *Evil and the Evidence for God: The Challenge of John Hick's Theodicy* (Philadelphia: Temple University Press, 1993). See pp. 439-40 of my review of Geivett's book in *Faith and Philosophy* 13 (1996) for a brief discussion of this.

24. See references in note 3.

25. And below I object to P2 - the inference from (II) to (III).

26. Thanks to Alvin Plantinga for reminding me of this.

27. The objection that follows was alluded to at the end of section 3. There I pointed out (roughly) that once we see that EP1\* entails EP1, my objection to the revised FWD entails the following sort of objection to the original FWD.

28. My thanks to Daniel Howard-Snyder, Hud Hudson, Trenton Merricks, Alvin Plantinga, Michael Rea, William Rowe and William Wainwright for helpful comments on earlier drafts. Special thanks are due to William Rowe for lengthy discussions out of which the idea for this paper grew.