

## *Freedom, Foreknowledge, and Frankfurt: A Reply to Vihvelin*

JOHN M. FISCHER  
University of California at Riverside  
Riverside, CA 92521  
USA

### **I Introduction**

In a fascinating and challenging article in this journal, Kadri Vihvelin presents a spirited and vigorous critique of the strategy of defending compatibilism about causal determinism and moral responsibility that employs (in part, at least) the 'Frankfurt-examples.'<sup>1</sup> Here is her presentation of such an example:

... Jones ... chooses to perform, and succeeds in performing, some action X. Tell the story so that it is vividly clear that Jones is morally responsible for doing X. If you are a libertarian, you may specify that Jones is an indeterministic agent who can choose otherwise, given the actual past and the laws. If you are a compatibilist, you may fill in the details so that Jones does X in a way that satisfies your favorite account of the counterfactual or dispositional facts that make it true that Jones could have done otherwise in the sense you think relevant to responsibility. Now, add to your story the following facts: there is standing in the wings another agent, Black. Black is interested in what Jones does. In particular, he wants Jones to do X and, moreover, Black has it in his power to prevent Jones from doing anything other than X.

The addition of Black to the story means that Jones could not have done other than X. But, Frankfurt argued, Jones is still responsible for doing X. After all, though

---

1 Kadri Vihvelin, 'Freedom, Foreknowledge, and the Principle of Alternate Possibilities,' *Canadian Journal of Philosophy* 8 (2000) 1-24

Black could have intervened, he didn't. He didn't have to. Jones chose to do X and did X without any interference from Black. So the addition of Black to our story doesn't remove or in any way diminish Jones's responsibility for doing X.

Such is the recipe for telling a Frankfurt story.<sup>2</sup>

Vihvelin is a vigorous critic of the Frankfurt stories.<sup>3</sup> Vihvelin's claims are vehement, and her argument intriguing. In light of these facts and also because I believe the issues are important, I wish to explore and critically evaluate her argument.<sup>4</sup> I shall begin by laying out the skeletal structure of the argument, after which I shall offer some critical ruminations.

---

2 'Freedom, Foreknowledge, and the Principle of Alternate Possibilities,' 4

3 Among other strongly critical statements, Vihvelin says:

If Frankfurt's aim was to convince libertarians that even if determinism renders us unable to do otherwise, it does not undermine responsibility, he has failed. If his aim was to make it easier to defend compatibilism, he has failed. And if his aim was to bypass questions about the truth-conditions of 'can do otherwise' claims, he has also failed, for the debate that has arisen in the wake of his original thought experiment is now mired deep in the very metaphysical questions he sought to avoid.

... It is my view that this literature is a philosophical dead end. Although I am a compatibilist, I think that Frankfurt's strategy for defending compatibilism is a bad one. If we begin with the commonsense view that someone is morally responsible only if she could have done otherwise, then Frankfurt stories will not and *should not* change our minds. If we are persuaded by Frankfurt, it is because we have been taken in by a bad argument ('Freedom, Foreknowledge, and the Principle of Alternate Possibilities,' 2-3).

4 In a recent post on the blog, 'The Garden of Forking Paths,' (<http://gfp.typepad.com>), Terrance Tomkow states:

It seems to me that Kadri Vihvelin (USC) demonstrated some time ago (CJP 2000) that Frankfurt's arguments turn on a modal fallacy. No matter how you tell these stories, Frankfurt Style Cases simply fail to describe agents who 'cannot do otherwise.'

If that's so then FSC cases have nothing to tell us about free will and the whole of the debate about FSCs has been a snare and a delusion.

To my knowledge no one has answered Vihvelin's arguments. So my question is: does anyone have an answer to Vihvelin or does the Frankfurt literature just keep stumbling forward out of inertia? ([http://gfp.typepad.com/the\\_garden\\_of\\_forking\\_pat/2005/06/compatibilism\\_a.html#comments](http://gfp.typepad.com/the_garden_of_forking_pat/2005/06/compatibilism_a.html#comments))

## II Vihvelin's Critique of Frankfurt-Style Compatibilism

The Frankfurt examples purport to show that an agent can be morally responsible for his actions (and even choices), even if he could not have done otherwise.<sup>5</sup> The examples actually go back (in some form or another) to John Locke, who argued that a man could voluntarily stay in a room, even though, unbeknownst to the man, the door to the room is locked.<sup>6</sup> The examples have been highly contentious, and their analysis and significance have generated a huge literature.<sup>7</sup> Some of us have agreed with Frankfurt's claim that moral responsibility does not require alternative possibilities, and additionally we have suggested that this fact can help in an overall argument (an argument that employs other ingredients as well) for the compatibility of causal determinism and moral responsibility.<sup>8</sup> My contention has been that, even though the arguments fall short of being decisive, there are strong plausibility arguments for the conclusion that the Frankfurt examples show that moral responsibility does not require alternative possibilities. Further, I have contended that the elements of the *direct* arguments for the incompatibility of causal determination and moral responsibility are considerably weaker than the ingredients of the arguments for the incompatibility of causal determinism and genuine metaphysical access to alternative possibilities (and thus the *indirect* arguments for the incompatibility of causal determinism and moral responsibility).<sup>9</sup>

---

5 Harry G. Frankfurt, 'Alternative Possibilities and Moral Responsibility,' *Journal of Philosophy* **66** (1969) 829-39

6 John Locke, *An Essay Concerning Human Understanding*, bk. 2, chap. 21, sec. 10.

7 For some of this literature, see John Martin Fischer, 'Some Recent Work on Moral Responsibility,' *Ethics* **111** (1999) 99-137; and David Widerker and Michael McKenna, eds., *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities* (Williston, VT: Ashgate 2003), 235-50.

8 John Martin Fischer, *The Metaphysics of Free Will: An Essay on Control* (Oxford: Blackwell Publishers 1994); 'Frankfurt-Type Compatibilism,' in S. Buss and L. Overton, eds., *Contours of Agency: Essays on Themes from Harry Frankfurt* (Cambridge, MA: The MIT Press 2002) 1-26, reprinted in G. Watson, ed., *Free Will: Oxford Readings in Philosophy*, 2nd Ed. (Oxford: Oxford University Press 2003) 190-211. and John Martin Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*.

9 For an elaboration and discussion of the two distinct kinds of arguments (direct and indirect), see John Martin Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*.

Vihvelin believes that all of the participants in the debates about the Frankfurt examples have failed to see some fundamental distinctions and logical problems. She highlights this view as follows:

I think we should have avoided this mess. Things went wrong from the start. No one should ever have been persuaded by Frankfurt's argument.<sup>10</sup>

Vihvelin begins her diagnosis of the errors of those of us who have invoked the Frankfurt examples by drawing a distinction between 'two ways of getting someone to do what you want':

Suppose you want to ensure that someone does whatever you want him to do, but, like Black, prefer to avoid showing your hand unnecessarily. There are two different methods you might employ.... I will call these 'the method of conditional intervention' and 'the method of counterfactual intervention.'

What makes someone a *conditional intervener* is the fact that his intervention is causally triggered by the beginnings of any action (overt or mental) contrary to the intervener's plan. If the subject begins to try or begins to do any undesired action, the intervener will prevent him from succeeding.

... What makes someone a *counterfactual intervener* is the fact that his intervention is causally triggered, *not* by the subject's trying or beginning to act contrary to the intervener's plan, but by some *earlier* event that is a reliable indicator of the fact that the subject will, in the absence of intervention, choose or act contrary to the intervener's wishes. This earlier event might be a blush, twitch, or other involuntary sign that occurs just before the subject begins to make an unwanted decision...<sup>11</sup>

Vihvelin goes on to state that this distinction has not been explicitly recognized in the literature about the Frankfurt examples, and that understanding it is 'the key to understanding both the seductive charm of Frankfurt stories and why they ultimately fail.'<sup>12</sup> Here her position is that mere conditional intervention cannot show that moral responsibility does not require *any* alternative possibilities (including freedom to choose otherwise), whereas counterfactual intervention can be seen to be problematic (logically defective). That is, it cannot 'in principle' work to show that moral responsibility does not require any alternative possibilities.<sup>13</sup>

Vihvelin's strategy for arguing that counterfactual intervention 'cannot in principle work' involves telling a certain story, evaluating it, and

---

10 'Freedom, Foreknowledge, and the Principle of Alternate Possibilities,' 8

11 'Freedom, Foreknowledge, and the Principle of Alternate Possibilities,' 9

12 'Freedom, Foreknowledge, and the Principle of Alternate Possibilities,' 11

13 'Freedom, Foreknowledge, and the Principle of Alternate Possibilities,' 11

then claiming that Frankfurt stories are in the relevant ways parallel. It will be helpful to have Vihvelin's presentation in some detail:

In my story you and I make a bet on the outcome of the toss of a coin. I bet heads — I always bet heads; you bet tails. It comes up heads. I win. You pay up. The question is: Did I win fairly? Well, of course it depends on whether it was a fair coin and a fair toss. Let's stipulate that it was. That is, let's stipulate that there was nothing about the physics of the coin or its toss that made it more likely that the coin would come up heads rather than tails. The odds were 50/50 that this toss would come up heads, and in the course of the toss no outside forces intervened to change those odds.<sup>14</sup>

Now Vihvelin adds that she has a 'confederate' named Black who can somehow or other (it is not known how) always predict the coin toss correctly. She goes on as follows:

Thanks to his unusual predictive powers, Black also has the ability to act ahead of time in a way that ensures that the coin will come up the way he wants it to. Here's how he does it. Black makes his predictions in the morning; we never toss the coin until evening. This leaves Black with plenty of time to fix either the coin or its environment, if need be, in a way that nomologically guarantees that the coin will come up the way that he wants. But Black also prefers not to show his hand unless he has to. If he predicts that the coin will come up the way he wants, then he does nothing. Either way, his job is done by noon. By the time we toss the coin, Black has retired for the day. You may imagine him far away, or fast asleep.

Finally, let's stipulate that Black is a friend of mine. He wants me to win. If he had predicted that without his intervention this particular coin toss would have come up tails, he would have fixed things so that it would have been nomologically necessary on this toss that the coin would come up heads.

But as a matter of fact, on that last toss, Black didn't intervene. He predicted (somehow) that the coin was going to come up heads, so he did nothing.<sup>15</sup>

Vihvelin maintains that, even in the scenario as elaborated in this way, she has won 'fair and square.' Her view is that, although it is true that the coin *will* never (given Black's powers) come up heads, nevertheless, given that Black does not intervene, the coin *could have* come up heads. She says that the 'complicated truth' about any given coin toss, under the circumstances described in the story, as filled in above, is:

---

14 'Freedom, Foreknowledge, and the Principle of Alternative Possibilities,' 14-15

15 'Freedom, Foreknowledge, and the Principle of Alternate Possibilities,' 15-16

EITHER the coin comes up heads even though it could have come up tails OR the coin comes up heads and could not have come up tails.<sup>16</sup>

Vihvelin goes on to say:

It follows from this that the coin *will* never come up tails, not that it *can't*. On some tosses the coin can [come] up tails as easily as it can come up heads. But, thanks to the peculiar setup, it so happens that the coin *can* come up tails only on those occasions that it actually comes up heads.<sup>17</sup>

Vihvelin's view, in short, is that the game is not rigged, when Black does not intervene, although it is rigged, when he does.

Finally, Vihvelin asserts that her story is parallel in relevant aspects to the Frankfurt stories. She says that both the libertarian and the compatibilist must say that the counterfactual intervener in a Frankfurt story does not rob the agent of the ability to choose otherwise.<sup>18</sup>

Vihvelin ends her article by suggesting that those of us who contend that an agent such as Jones is morally responsible but cannot choose or do otherwise are guilty of a logical blunder:

More austere, the point is this. The inference from:

(P and Possibly not-P) or (P and Necessarily P)

To

Necessarily P

is fallacious. To suppose otherwise is to permit the inference from:

P

To

Necessarily P

And this is the logic of the fatalist.<sup>19</sup>

---

16 'Freedom, Foreknowledge, and the Principle of Alternate Possibilities,' 18

17 'Freedom, Foreknowledge, and the Principle of Alternate Possibilities,' 18

18 'Freedom, Foreknowledge, and the Principle of Alternate Possibilities,' 22-3. Vihvelin goes on to assert that a 'complicated truth' — parallel to the complicated truth about her story — obtains here.

19 'Freedom, Foreknowledge, and the Principle of Alternate Possibilities,' 23

### III Reply to Vihvelin

I wish to commend Vihvelin for her subtle, probing, and suggestive critique. It will be helpful to analyze Vihvelin's critique carefully. Before delving into her 'story,' I wish to express my agreement with Vihvelin that there is indeed an important distinction between 'purely conditional interveners' and 'counterfactual interveners.' Further, I agree that a purely conditional intervener cannot be invoked to show (without further argumentation) that moral responsibility does not require *any* alternative possibilities. Vihvelin and I are in agreement that the focus should be on counterfactual interveners, rather than purely conditional interveners.

Turn back to Vihvelin's story. Note, again, the way she describes her Black. Among other things, she says, 'Black is able, somehow, to predict how any given coin toss will turn out if he does not intervene and his predictions are *always* right, no matter how fair the coin and its environment.'<sup>20</sup> It is unclear what the italics add to 'always,' but they clearly play some rhetorical role at least. Perhaps the italics help to suggest that Black somehow has *knowledge* of the future, although Vihvelin backs off from such a suggestion in footnote 23.<sup>21</sup> For my purposes, I could grant that somehow or other Black has genuine knowledge of the future; nothing in my reply to Vihvelin depends on whether or not Black has knowledge (as opposed to belief with a high degree of justification or confidence).

Suppose that Black does indeed know in advance how the coin will come up later in the day. He then 'retires for the day,' well before the coin is flipped. My claim is that, given that Black has indeed retired for the day — he is 'asleep or far away' — the following counterfactual is true (at the relevant time — just prior to the coin toss): if the coin were about to come up tails, Black would not be able to intervene and thus the coin would in fact come up tails. Given the set-up of the situation, even assuming Black's knowledge in the morning, if the coin were about to turn up tails, it would do so; under this scenario, although Black actually knows how the future will come up, he would not have known (in the alternative sequence). That is, he would have had a false belief in the scenario in which the coin turns up tails. (Again, nothing depends on supposing Black actually has knowledge — we could equally well suppose that he simply has confident belief, a belief that would have turned out false in the alternative sequence.)

---

20 'Freedom, Foreknowledge, and the Principle of Alternate Possibilities,' 15

21 'Freedom, Foreknowledge, and the Principle of Alternate Possibilities,' 15

My diagnosis of the flaw in Vihvelin's argument is that she posits that her story is parallel to Frankfurt stories, but it is in fact crucially different: the relevant counterfactuals are *not* true in her story, whereas they *are* true in the Frankfurt examples. In virtue of this lack of parallelism in the counterfactuals, the modal facts in Vihvelin's story and the Frankfurt examples are crucially different. That is, whereas it is not true in Vihvelin's story that the coin could not have turned up tails, it seems to me that in the Frankfurt examples the relevant agent lacks the power to choose and do otherwise.

Note that Vihvelin has not supposed that Black is somehow essentially infallible; that is, Black is not alleged to be such that he has as part of his 'essence' that he has all and only true beliefs.<sup>22</sup> Under such an assumption, it would arguably be the case that the relevant counterfactual would be true; that is, it would be true that if the coin were about to come up tails, then Black would have known this and would have taken precautions to ensure that it would not have succeeded in coming up tails. But whereas the assumption of Black's essential infallibility would arguably underwrite the truth of the sorts of counterfactuals that are true in the Frankfurt cases, Vihvelin does not make such an assumption. Further, this assumption would render it completely contentious that the coin could have come up tails, as all of the intractable debates about the relationship between essential foreknowledge and 'possibility' would be engaged.<sup>23</sup>

I wish to elaborate here. It is generally conceded that there is a crucial difference between mere human foreknowledge and (say) divine foreknowledge. If some human being genuinely knows a contingent proposition about the future, then it follows that the proposition will turn out to be true; this is because knowledge implies truth of the proposition known. But it is widely recognized that it does *not* follow from the fact that some human being knows a contingent proposition about the future that the proposition is *necessarily* true (in the relevant sort of necessity), or that the events described or contained in the proposition *must* take place (in the relevant sense of 'must'). Although the *conditional* 'If *S* knows that *p*, then *p* is true' is necessary, this necessity does not attach to the consequent of the conditional, even given the truth of the antecedent. (This of course reflects the well-known distinction

---

22 For discussions of the assumption of essential infallibility, see John Martin Fischer, ed., *God, Foreknowledge, and Freedom* (Stanford: Stanford University Press 1989).

23 John Martin Fischer, ed., *God, Foreknowledge, and Freedom* and John Martin Fischer, 'Recent Work on God and Freedom,' *American Philosophical Quarterly* 29 (1992) 91-109

between the ‘necessity of the consequence’ and the ‘necessity of the consequent.’)

So even if we assume that in Vihvelin’s story Black is able to predict a later coin toss in such a way as genuinely to know how the coin will come up, it does not follow merely from this fact that later the coin cannot come up differently. If the coin flip had come up differently, then Black would have had a false belief — he would not have known something he actually did know. In contrast, if an essentially omniscient agent — say God — knows how a later coin toss will turn out, then the only way it could have come up differently would be if the past had been different from what it actually was; that is, the only way it could have come up differently would be if God had had a different belief in the past. But since the past is in relevant ways ‘fixed’ and ‘over-and-done-with,’ it is not clear that, given the assumption of God’s essential omniscience, the coin *could* have come up differently. (This is obviously an analogue of the problem of the apparent incompatibility of God’s foreknowledge and human freedom, in the sense that requires ‘freedom to do otherwise.’)

The key point is that nothing in Vihvelin’s story (as she tells the story — and without an assumption of Black’s essential omniscience) plausibly rules it out that, although Black knows how the coin toss will come out, the coin can come up on a different side. Presumably Black could still ensure that the coin come up heads, if he were to stay on the scene and monitor the situation. But, according to Vihvelin, he does *not* do this: ‘he retires for the day,’ and is thus not in a position to ensure that the coin comes up heads. A babysitter might well be in a position to ensure that a young child stay in his room by standing outside the door and monitoring the situation; if the child were to try to come out the door, the babysitter can stop him from doing so. But if the babysitter decides to watch television in another part of the house, or just goes home (every parent’s nightmare — or one of them!), then the babysitter is obviously no longer in a position to prevent the child from leaving the room.

In summary, the problem then with Vihvelin’s argument is that it posits that her stories are parallel to the Frankfurt stories, but they differ with regard to the crucial counterfactuals. In the coin case, if the coin were about to come up tails, it would — Black’s prediction would have been false, and he wouldn’t be there to stop it. In the Frankfurt case, if Jones were about to refrain, he would not, because Black is there to intervene. I claim that, in virtue of the truth of the relevant counterfactuals in the Frankfurt cases, the agent in those cases lacks the freedom to choose or do otherwise. So even if it is granted that the coin could come up tails in Vihvelin’s story, this would not show anything about the Frankfurt cases. And if Vihvelin were to strengthen her assumption

about Black, making him essentially omniscient, then she could not safely claim that the coin could have come up tails! Thus she couldn't argue from a parallel with her story to her desired conclusion about the Frankfurt cases.

It should now be evident that I would clearly *not* argue in the fatalistic way suggested at the end of Vihvelin's paper. That is, I would not begin with:

(P and Possibly not-P) or (P and Necessarily P).

Rather, the analysis I favor would characterize the Frankfurt cases in terms of the second disjunct (together with the suggestion that the way in which the 'necessity' in question is achieved is compatible with moral responsibility in the actual sequence).

#### IV The Logic of the Frankfurt-Style Argument (Continued)

Not only does Vihvelin contend that a proponent of certain views about the Frankfurt cases (according to which the agent is morally responsible but does not have genuine access to alternative possibilities) is implicitly relying on problematic fatalistic modes of reasoning, but she also claims that reliance on the sorts of counterfactuals to which I have pointed above is similarly logically suspect. Vihvelin finds such reasoning in a paper I wrote with Paul Hoffman:

If he were about to refrain (in the absence of intervention by an external agent or factor) the triggering event would already have occurred. [If the triggering event had already occurred, Black would have intervened and forced Jones to act, in which case Jones would not have been able to refrain.] If Jones were about to refrain, he would be rendered unable to refrain.<sup>24</sup>

But Vihvelin rejects this form of argumentation, saying:

... this ... relies on a form of counterfactual reasoning which is generally considered invalid: hypothetical syllogism. An example: 'If I jumped off this bridge, I would have arranged to be wearing a parachute. If I were wearing a parachute, I would not be killed. So if I jumped off this bridge, I would not be killed.'<sup>25</sup>

---

24 The passage is from John Martin Fischer and Paul Hoffman, 'Alternative Possibilities: A Reply to Lamb,' *Journal of Philosophy* 90 (1993) 517-27. It is quoted by Vihvelin on p. 20 (n. 30) of 'Freedom, Foreknowledge, and the Principle of Alternate Possibilities.'

25 'Freedom, Foreknowledge, and the Principle of Alternate Possibilities,' p. 20.

I reply that no proponent of the conclusion — based on consideration of the Frankfurt examples — that moral responsibility does not require alternative possibilities of whom I am aware has ever proceeded by supposing that he or she could construct an argument that employs *solely* the form of reasoning identified by Vihvelin. In particular, I have never thought that one could simply argue for my favored conclusions about the Frankfurt examples by employing the form of reasoning (hypothetical syllogism involving counterfactuals), on the supposition that this is a formally valid way of reasoning. Rather, my point is that the conclusion in question follows from the relevant premises *together with other facts*. These facts about the examples are precisely the sort that help to license an inference to the relevant conclusion, even though the inference form in question is not *formally valid*.

To explain this point, consider David Lewis's remarks in his classic discussion of these matters:

For a direct counterexample to transitivity [and presumably hypothetical syllogism], consider this argument:

If Otto had gone to the party, then Anna would have gone.

If Anna had gone, then Waldo would have gone.

Therefore: If Otto had gone, then Waldo would have gone.

The fact is that Otto is Waldo's successful rival for Anna's affections. Waldo still tags around after Anna, but never runs the risk of meeting Otto. Otto was locked up at the time of the party, so that his going to it is a far-fetched supposition; but Anna almost did go. Then the premises are true and the conclusion false. Or take this counterexample, from Stalnaker:

If J. Edgar Hoover had been born a Russian, then he would have been a communist.

If he had been a Communist, he would have been a traitor.

Therefore: If he had been born a Russian, he would have been a traitor.

In general, transitivity [and presumably hypothetical syllogism] fails... [when] the antecedent of the first premise [is] more far-fetched than the antecedent of the second, which is the consequent of the first. Then the closest worlds where the first antecedent holds are different from — and may differ in character from — the closest worlds where the second antecedent holds.<sup>26</sup>

---

Vihvelin here refers to various discussions of the alleged 'fallacy' in question: P. J. Downing, 'Subjunctive Conditionals, Time Order, and Causation,' *Proceedings of the Aristotelian Society* 59 (1958-59), 125-40; David Lewis, *Counterfactuals* (Cambridge, MA: Harvard University Press 1973), 31-6; and Jonathan Bennett, 'Counterfactuals and Temporal Direction,' *Philosophical Review* 93 (1984), 57-91.

Lewis goes on to characterize further the situations in which the relevant form of inference fails, and how to strengthen the premises to avoid a failure of inference.<sup>27</sup>

On the approach to counterfactual conditionals defended by Lewis (and Stalnaker), a counterfactual such as 'If *P* had been the case, then *Q* would have been the case' is true (roughly speaking) just in case *Q* is true in the possible world or worlds in which *P* is true that is (or are) 'closest' to the actual world (is — or are — the 'most similar to' or 'involve the least large departure from' the actual world). What is crucial for my discussion here is that transitivity and hypothetical syllogism fail in cases with a specific structure: as Lewis puts it, the antecedent of the first premise must be more far-fetched than the antecedent of the second, which is the consequent of the first. This issues in the possibility that the possible worlds relevant to the truth of the first premise are *different from* the possible worlds relevant to the truth of the second premise. This sort of 'world-hopping' is the semantic basis for the possibility of the failure of the relevant forms of inference: the various premises are 'sending us to different possible worlds.'

Of course, it is important to ask whether anything like this is happening in the Frankfurt-examples and in my (or others') arguments that employ these examples. It seems clear to me that nothing like the structure identified by Lewis is present in the Frankfurt stories or my (or anyone's) analysis of them. As stated above, Vihvelin's way of regimenting the argument is as follows:

If [the relevant agent] were about to refrain (in the absence of intervention by an external agent or factor), the triggering event would already have occurred.

If the triggering event had already occurred, Black would have intervened and forced Jones to act, in which case Jones would not have been able to refrain.

Therefore: If Jones were about to refrain, he would be rendered unable to refrain.

As far as I can see, there is no problematic 'world-hopping' in this example or the argument that accompanies it (as regimented just above). There is just one world (or set of worlds) to which the premises send us — it is not the case that the antecedent of the first premise is more far-fetched than the antecedent of the second (which is the consequent of the first). Indeed, the story is supposed to be a coherent story about one possible scenario. Hence, it would seem that the premises imply the conclusion, given the facts about the example, even if transitivity-

---

27 David Lewis, *Counterfactuals*, 33-6

ty and hypothetical syllogism are formally invalid for counterfactual conditionals.<sup>28</sup>

Because the issues are somewhat delicate, I wish to seek to explain my view here a bit more fully. Note, again, the sort of ‘world-hopping’ that is taking place in Stalnaker’s example discussed above:

If J. Edgar Hoover had been born a Russian, then he would have been a communist.

If he had been a Communist, he would have been a traitor.

Therefore: If he had been born a Russian, he would have been a traitor.

The first premise is true in virtue of a possible world in which Hoover is a Russian; but the second premise is true in virtue of a world in which Hoover is not a Russian. In going from the first to the second premise, we have hopped from one world (or set of worlds) to another. Therefore, there is no guarantee that there is a possible world in virtue of which the conclusion is true.

The same sort of world-hopping takes place in Lewis’s example:

If Otto had gone to the party, then Anna would have gone.

If Anna had gone, then Waldo would have gone.

Therefore: If Otto had gone, then Waldo would have gone.

The first premise is true in virtue of a world in which Otto has somehow escaped the locked room of the actual world; but the second premise is true in virtue of a world in which Otto is still locked up. Therefore, as above, there is no guarantee that there is a possible world in virtue of which the conclusion is true.

---

28 For a similar point about a different argument, see: John Martin Fischer, ‘A New Compatibilism,’ *Philosophical Topics* 24 (1998) 49-66; reprinted in Laura Waddell Ekstrom, ed., *Essays on the Metaphysics of Freedom* (Boulder, CO: Westview Press 2001) 38-56. Here I respond to a criticism that a version of an argument for the incompatibility of causal determinism and freedom to do otherwise I sketched is ‘formally invalid.’ I contend that the argument is valid, given its form *and the content of the premises*. For further discussion and defense, see: John M. Fischer and Mark Ravizza, ‘When the Will is Free,’ *Philosophical Perspectives* 6 (1992) 423-52, reprinted in Timothy O’Connor, ed., *Agents, Causes, and Events* (Oxford: Oxford University Press 1995) 239-69; and John Martin Fischer and Mark Ravizza, ‘Free Will and the Modal Principle,’ *Philosophical Studies* 83 (1996) 213-30. Also, see John Martin Fischer, ‘Critical Notice of J. Howard Sobel’s *Puzzles of the Will*,’ *Canadian Journal of Philosophy* 31 (2001) 427-44.

But now consider again the reasoning that Vihvelin supposes is behind my analysis of the Frankfurt-type case:

If [the relevant agent] were about to refrain (in the absence of intervention by an external agent or factor), the triggering event would already have occurred.

If the triggering event had already occurred, Black would have intervened and forced Jones to act, in which case Jones would not have been able to refrain.

Therefore: If Jones were about to refrain, he would be rendered unable to refrain.

Given the story of the Frankfurt-type case, I do not see any reason to suppose that a structurally similar sort of world-hopping is taking place here. As far as I can see, the Frankfurt-story posits a single possible scenario in virtue of which the two premises are true. Thus, there is no reason to suppose that there is no possible world in virtue of which the conclusion is true; it is precisely the same single world in virtue of which the premises are true.

To develop the point further, consider the form of reasoning under consideration (transitivity or hypothetical syllogism for counterfactual conditionals):

If P had been the case, then Q would have been the case.

If Q had been the case, then R would have been the case.

Therefore: If P had been the case, then R would have been the case.

David Lewis points out that in all counterexamples to the form of inference in question, the 'might-counterfactual,' 'If Q had been the case, then P might not have been the case,' is true.<sup>29</sup> He also points out that in some of the counterexamples (and, in particular, in the Otto/Waldo case and the J. Edgar Hoover case) the 'would-counterfactual,' 'If Q had been the case, then it would still not have been the case that P,' is non-vacuously true.<sup>30</sup> So, for example, in Lewis's words, 'If Anna had gone, Otto would still not have; if Hoover had been a Communist, he would still not have been born a Russian.'<sup>31</sup> Lewis goes on to explain how 'adding a third premise to the inference by transitivity, we may rule out all cases where transitivity [hypothetical syllogism] fails.'<sup>32</sup>

---

29 David Lewis, *Counterfactuals*, 33

30 David Lewis, *Counterfactuals*, 33

31 David Lewis, *Counterfactuals*, 33

32 David Lewis, *Counterfactuals*, 33-4

For my purposes here the key point is to note that in the Frankfurt stories, the characteristic structure of a counterexample to the relevant form of inference is *absent*. In order for a Frankfurt story to be a candidate for being such a counterexample, the following ‘might-counterfactual’ would need to be true: ‘If the triggering event had already occurred, then it might not have been the case that [the relevant agent] was about to refrain (in the absence of intervention by an external agent or factor).’ But the Frankfurt stories include the fact that the ‘counterfactual intervener’ is a reliable predictor of the agent’s future choices and behavior, and the basis for this reliability is the ‘prior sign’ or ‘triggering event’ (in some Frankfurt stories, at least). That is, in the Frankfurt stories it is supposed that the prior sign or triggering event would occur if and only if the agent were about to choose or do otherwise.<sup>33</sup> So the specific feature identified by Lewis as common to all counterexamples to the inference-form is absent from the Frankfurt cases!

The upshot of the discussion above is that, if one regiment my analysis of the Frankfurt cases as suggested by Vihvelin, this regimentation lacks the characteristic structure of counterexamples to the rule of inference it employs. Additionally, I am not at all convinced that one *needs* to regiment the analysis as suggested by Vihvelin. Here are the facts in a Frankfurt case. There is a triggering event that occurs to indicate that the agent is about to refrain. The ‘counterfactual intervener’ watches for that and even has the power and intention to intervene upon noticing the triggering event. Further, the counterfactual intervener is a completely reliable triggering-event detector, and is completely reliable in carrying out his intentions. Given these facts, it just seems intuitively obvious that if the relevant individual (Jones) were about to refrain, he would be rendered unable to refrain. And thus it seems intuitively obvious that Jones is unable to do otherwise, given the facts of the case;

---

33 There has been considerable controversy about whether the Frankfurt cases are really cases in which this sort of biconditional is true; but Vihvelin believes we can avoid getting into the details of these discussions entirely. This is important to keep in mind when evaluating Vihvelin’s critique, which she supposes is more fundamental than others, and can help us to side-step the complex and voluminous literature on Frankfurt cases. To be a bit more explicit, some might call into question some version of the counterfactual, ‘If [the relevant agent] were about to refrain (in the absence of intervention by an external agent or factor), the triggering event would already have occurred.’ It might be considered contentious, within the dialectic of the debates between compatibilism and incompatibilism, whether such a counterfactual is true. But I wish to avoid such controversies here; after all, Vihvelin has provided no argument that such a counterfactual cannot be true, and her contention is that employing her critique can help us to see a deeper, more general problem with Frankfurt examples.

no argument employing hypothetical syllogism or transitivity appears to be required.<sup>34</sup>

## V Conclusion

Kadri Vihvelin has issued an important challenge to those of us who have sought to argue that Harry Frankfurt's famous examples help to show that moral responsibility does not require genuine access to alternative possibilities. She does not focus her attention on the arcane debates in the literature about sophisticated versions of the examples or their proper analysis; rather, she contends that we are all missing some basic logical facts.

I have tried to address Vihvelin's challenge here. I have pointed out that in her 'story,' the relevant counterfactuals — those allegedly parallel to the relevant counterfactuals in the Frankfurt stories — are not true; it is thus natural to suppose that the modal fact which (according to many of us) obtains in the Frankfurt stories does not obtain in Vihvelin's story. Although she wishes to argue from a putative parallelism between her story and the Frankfurt examples, there is a crucial asymmetry. Further, there is no basic logical fallacy in arguing, as many of us have, that the set-up of the Frankfurt cases implies that, although the agent in question acts freely and is morally responsible, he lacks the relevant sort of access to alternative possibilities.<sup>35</sup>

*Received: October 2005*

*Revised: June 2006*

---

34 Of course, the discussion in the text above shows that the analysis can indeed be regimented in the way suggested by Vihvelin without rendering it invalid, given that it is understood in light of the comments by David Lewis about strengthening the premises.

35 I wish to thank Neal A. Tognazzini and four anonymous referees for the *Canadian Journal of Philosophy* for their helpful comments.