Foreknowledge, Frankfurt, and Ability to Do Otherwise: Reply to Fischer

There is one important point about which Fischer and I are in agreement. We agree that determinism is compatible with moral responsibility. We disagree about the best way of defending that claim. He thinks that Frankfurt's strategy is a good one, that we can grant incompatibilists the metaphysical victory (that is, agree with them that determinism means that we are never able to do otherwise) while insisting that we are still morally responsible. I think this a huge mistake and I think the literature spawned by Frankfurt's attempt to undercut the metaphysical debate between compatibilists and incompatibilists is a snare and a delusion, distracting our attention from the important issues.

I am a compatibilist, not a "semi-compatibilist" but an unabashed traditional compatibilist who believes that determinism (by itself) doesn't render us unable to do otherwise in any significant or morally relevant sense. I don't make these claims lightly; I think that the problem of free will and determinism is a metaphysical problem which can be solved only by paying careful attention to modal and metaphysical issues concerning choice, agency, ability, dispositions, counterfactuals, causation, laws, and so on. I have defended this view in a number of places[1]. In my "Freedom, Foreknowledge, and The Principle of Alternate Possibilities" [2] I argued that so-called "Frankfurt stories" *cannot* show what they are supposed to show: that a person who is never able to do otherwise may nevertheless be morally responsible for what she does.[3] I offered a diagnosis of why they have persuaded so many for so long. It has gone unnoticed that there are two very different methods that Black might use to ensure that Jones does exactly what Black wants him to do; Black might be what I called a 'conditional

intervener' or Black might be what I called a 'counterfactual intervener'. Both interveners have genuine and powerful powers, and it is natural to suppose that if we grant Black *both* kinds of powers, then Black succeeds in depriving Jones of all his morally significant alternatives. But if we carefully examine the two methods, we will see that while the powers of a conditional intervener are genuine, they are also limited. If Black is *only* a conditional intervener, then he cannot, in principle, rob Jones of the kind of freedom traditionally regarded as essential to free will and moral responsibility: the freedom to choose or at least to try or begin to choose otherwise. At first glance, it appears that if we grant Black the powers of a counterfactual intervener, then Black can ensure that Jones lacks even this inner freedom. But, I argued, if Black is *only* a counterfactual intervener, he does not succeed in robbing Jones of any freedom whatsoever. If we think that he does, it is because we have let ourselves be persuaded by the intuitions or arguments that underlie the fatalist's conflation of truth and necessity. In passing, I also pointed out that one way of defending the claim that Jones is unable to do otherwise – a so-called "back-tracking" argument -- relies on hypothetical syllogism, which is generally considered invalid for counterfactual conditionals.[4]

Fischer's response[5], in this journal, can be summed up as follows:

1. He agrees that the distinction between conditional and counterfactual intervention is an important one, and that a purely conditional intervener cannot be invoked to show that moral responsibility does not require *any* "alternative possibilities". He agrees that "the focus should be on counterfactual interveners".
2. He agrees that my story about Black and the coin succeeds in showing that the mere *existence* of a counterfactual intervener does not suffice to show that a

person is unable to choose otherwise (or begin or try to choose otherwise). So he appears to concede that additional argument is needed.

3. He does not take issue with my criticism of the arguments which rely on variations of fatalist reasoning nor does he dispute my claim that hypothetical syllogism is invalid for counterfactuals.
4. He defends the "back-tracking" argument I criticised by arguing that the addition of further facts changes it into an argument that is valid and he claims that *in the right kind of Frankfurt story* the premises of the argument are true.
5. Finally, he suggests that given the right kind of Frankfurt story there is no need of an *argument* in defense of the claim that Jones lacks all morally significant alternatives because it is "intuitively obvious" that Jones is unable to do otherwise.

I will address Fischer's response to my paper in due course. But first, let me lay out the groundwork. For while I think that arguments based on Frankfurt stories go wrong, I don't think the mistakes are simple or obvious ones.

The Year That Jones Didn't Ride His Bike

Let's start with something simpler than the ability to do otherwise - the ability to ride one's bicycle. Suppose that Jones has free will (understood whatever way you like) and suppose that he knows how to ride a bike, and often does, then stops doing so. A year goes by, and Jones never rides his bike. He doesn't feel like it. We would ordinarily suppose (barring something like broken legs or brain damage) that Jones still has the ability to ride his bike. But I haven't told you the whole story. There was, all along, in the background but paying very close attention, a powerful figure named Black, with a mysterious interest in Jones's doings. Black had a plan for Jones and he was prepared to

intervene to enforce his plan, but only if absolutely necessary. Since Black is so powerful, he always gets his way. It was Black's plan, that year, that Jones not ride his bike. If Jones had ever tried to ride his bike, Black would have stopped him. But Jones never wanted to ride, so Black never intervened.

What difference does the addition of these extra facts – the facts about Black – make to our story about Jones? We should admit, I think, that Black succeeded in making a difference to *some of the modal facts* about Jones. Before Black came along, riding his bike was something that Jones could often do in a straightforward, ordinary sense of 'could'; he had the ability and he also had the opportunity. [6] Black didn't remove Jones's opportunities in any of the usual ways; he didn't break Jones' bike, or steal it, nor did he lock Jones up; all he ever did was pay very close attention to Jones. But insofar as Black had the power and the determination to carry out his plan for Jones, we should regard Black as someone who robbed Jones of opportunities. A prisoner in a cell with doors which lock automatically when the door is approached is as effectively constrained as a prisoner in a room with locked doors. By hanging about, ready to do whatever it took to prevent Jones from riding his bike, Black was like the conditionally locking doors of the prison cell: he deprived Jones of the *opportunity* to ride his bike.

On the other hand, since Black never actually laid a finger on Jones – he did not break his legs or mess with his brain – Jones' *ability* to ride his bicycle remained intact. Before Black came along, there were many occasions on which it was true that Jones had the ability to ride his bike; he knew how, he had the relevant skills, and he was physically and psychologically capable of exercising those skills -- his legs weren't broken, he wasn't suffering from loss of muscle control, pathological fear of bike-riding,

incapacitating depression, and so on. [7] Black did nothing to change *these facts*. YYou don't lose your ability to do something just because you don't exercise it; you also don't lose your ability just because someone prevents or would prevent you from exercising it. Jones retained the ability to ride his bike during the year that Black watched him; it remained true of him that he had "what it takes" to ride it.

It makes no difference how powerful Black is. Let's stipulate that Black has the power to thwart any and all of Jones's attempts to ride his bike and an unwavering resolution to use his power if (but only if) it becomes necessary. Let Black hover over Jones constantly, ready to intervene at lightning speed the instant he detects in Jones the first sign that Jones is trying or beginning or beginning to try to ride his bike. Let the following counterfactuals all be true: If Jones got on his bike and tried to ride it, Black would prevent Jones from succeeding (eg. by causing Jones to lose control over his legs). If Jones decided to ride his bike, Black would stop him before he even sets foot on the bike (eg, by pushing him away). If Jones began to approach his bike with the intention of riding it, Black would prevent him from reaching it (eg. by snatching his bike). And so on. The basic point remains the same. Insofar as it's true that Black would have acted in these kinds of ways to frustrate Jones' bike-riding decisions, plans, and efforts, Black robbed Jones of the *opportunity* to ride his bicycle. But Black did not rob Jones of his *ability* to ride his bicycle.[8]

1969: Frankfurt's Metaphysical Conjuring Trick

As we all know, Harry Frankfurt[9] changed the rules of the game for the debate between compatibilists and incompatibilists. Until 1969, the debate was about whether determinism deprives us of free will and moral responsibility by rendering us unable to

do otherwise. Frankfurt pointed out that both sides to the debate accepted an assumption that he unfortunately[10] named "The Principle of Alternate Possibilities" (**PAP**): A person is morally responsible for what he has done only if he could have done otherwise. The debate was about how "could have done otherwise" should be understood, but everyone agreed that **PAP** itself is true; indeed, many thought it an a priori truth. Frankfurt proposed to show that **PAP** is false, thereby undercutting the traditional debate. His basic idea was simple and brilliant. It took the form of a schema for a thought experiment that is supposed to show you that **PAP** is false, *regardless of your views about the proper understanding of "could have done otherwise".* By following his directions, we are supposed to see that we have been confused, all along, about what is necessary for moral responsibility. There are two steps to the thought experiment.

Step one: The set up. Tell a story about a person (we'll follow Frankfurt and call him "Jones") who makes a decision and performs an action and tell it so that it is vividly clear that Jones has free will and *can do otherwise in whatever sense you think is necessary for moral responsibility.* If you are an incompatibilist, you may specify that Jones lives at an indeterministic world and that he has agent-causal powers or whatever else you think is required. If you are a compatibilist you should fill in the story so that Jones satisfies your favorite account of the modal facts that you think make it true that Jones could have done otherwise.[11]

Step Two: The magic (in which Jones is rendered unable to do otherwise in a harmless, responsibility-preserving way). Add to your story a powerful person named Black, with a mysterious interest in Jones's doings. Black has the power to make Jones do whatever he wants. (Again, you may fill in the details so that when Black intervenes it

is vividly clear to you that Jones cannot do otherwise.) But Black prefers not to intervene unless he has to. Since Black is such a good predictor, he waits until *just before* Jones is about to make up his mind what to do.

> "He does nothing unless it is clear to him… 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' initial preferences and inclinations, then, Black will have his way." (p.6)

By a happy co-incidence, Jones makes the very decision that Black wants him to make, for his own reasons. So Black merely watches as Jones decides and does exactly what Black wanted him to decide and do.

The facts about Black make it true, Frankfurt tells us, that Jones could not have avoided making the decision that he made and performing the action that he performed. "What action he performs is not up to him." "He has no alternative but to do what Black wants him to do." [12] But because Jones acted for his own reasons, and not because Black forced him, Jones remains morally responsible.

It is easy to agree with Frankfurt about the claim that Jones remains morally responsible. The hard part is understanding how Black could have rendered Jones unable to do otherwise *without ever laying a finger on him.* That seems like magic. Of course, Black was prepared to intervene. And *if he had intervened*, then we can agree that Jones would have been unable to do otherwise. But Black didn't intervene. So it seems that there is *something* that is left up to Jones, something that he doesn't do but is able to do,

even if that something is only to begin or to try or to begin to try to make a decision contrary to Black's will. And so long as we are convinced of this, we will continue to defend **PAP**.

Frankfurt never pulled off his metaphysical conjuring trick. He left us with a promissory note, which his supporters have been trying to cash ever since.

Vihvelin 2000: Two Ways of Getting Someone to Do What you Want

In the paper[13] to which Fischer is replying, I argued that the literature dedicated to arguments about Frankfurt stories is a philosophical dead end. I offered a conceptual tool for helping us understand how Frankfurt's schema for a thought experiment has led us astray. I argued that there are two different kinds of powers or abilities Black might be thought to have, corresponding to two different ways in which he might be prepared to intervene. Black might be what I called a "conditional intervener"; if so, then Black's intervention would be causally triggered by the *beginnings of any action or attempted action (*including mental actions, like decisions) that is contrary to Black's plan for Jones. Or Black might be what I called a "counterfactual intervener"; if so, then Black's intervention would be causally triggered by some *earlier event* that is a *reliable indicator* of the fact that Jones will, in the absence of intervention, begin or try to perform some *action* (again, including mental actions) that is contrary to Black's plan for Jones. The difference between these two ways of intervening is not the coercive machinery employed at the non-actual worlds where Black intervenes: both interveners act in a way that satisfies Frankfurt's direction that you fill in the details so that it is clear to you that Black's *intervention* renders Jones unable to do otherwise. The difference, at these non-

actual worlds, lies in the timing of the intervention: the counterfactual intervener intervenes *earlier* – before Jones even begins or tries to make a decision or perform some other action contrary to Black's plan. And the difference, at the actual world where Black never intervenes, lies in the fact that *different causal counterfactuals are true.* If Black is a conditional intervener, then it's true, for every action X contrary to Black's plan for Jones, that if Jones had *tried* or *begun* to do X, this *would have caused* Black to prevent him from succeeding. If Black is a counterfactual intervener, then it's true, for ever action X contrary to Black's plan, that if some *earlier event* which is a reliable indicator of Jones' trying or beginning to do X had occurred, this *would have caused* Black to prevent Jones from even trying or beginning to do X.

In the Frankfurt stories typically told in the literature, Black is described as if he is *both* a conditional and a counterfactual intervener, and we are confused, dimly recognizing that a conditional intervener cannot (so long as he does not intervene) make it true that there is *nothing* that is up to Jones but thinking that a counterfactual intervener can somehow take up the slack. My argument, in a nutshell, is as follows: Black is either a conditional or a counterfactual intervener, or he employs a combination of the two ways of intervening. (There are no other possibilities.) We can achieve conceptual clarity by focusing on the "pure" cases where Black employs only one of the ways. If Black is a pure conditional intervener, than he does not succeed in robbing Jones of the kinds of morally relevant abilities we thought he had at the first stage of the thought experiment; in particular, Black does not rob Jones of the ability to make decisions, nor does he rob Jones of the ability to make a decision other than the one he actually makes. So Frankfurt's promissory note is cashed only if pure counterfactual intervention works;

that is, only if a Black who is *only* a counterfactual intervener succeeds in depriving Jones of the ability to decide (or begin or try to decide) otherwise. But if Black is a pure counterfactual intervener, he doesn't succeed in depriving Jones of *any* abilities, nor does he succeed in robbing him of any opportunities. If we think that he does, it is because we have fallen victim to fatalist thinking or some other kind of mistaken modal or counterfactual reasoning.

Black as Pure Conditional Intervener

The easiest way to understand the genuine power (as well as the limits) of conditional intervention is to revisit my story about the year that Jones didn't ride his bicycle. My original story, with Black hovering in the background, ready to intervene the instant he saw Jones beginning or trying to ride his bicycle, was a story about a conditional intervener. I pointed out that in that story, Black did succeed in changing some of the modal facts about Jones: Black robbed Jones of the opportunity to ride his bike that year. But Black did not succeed in depriving Jones of the ability to ride his bike.

We can make the story more like a Frankfurt story by supposing that Black's plan for Jones was much more detailed, encompassing every detail of Jones' doings that year, including his mental doings. In this version of the story Black paid close attention not only to Jones' body but also to his brain, and he was prepared to leap into action the instant he saw Jones beginning to do, decide, or even to think anything contrary to Black's plan. But Black never had to intervene because by happy co-incidence Jones always thought, decided, and acted in exactly the ways Black wanted him to.

My point remains unchanged. In this more elaborate version of the story, Black succeeded in robbing Jones of more opportunities, but he still did not deprive him of any

abilities. Thanks to Black's constant readiness to jump in to prevent Jones from straying from the plan, Jones was robbed of all opportunities for successful action except those that conformed to Black's plan.[14] Whereas in the simpler version of the story, Black was like the prison doors which lock automatically on the approach of the prisoner (robbing Jones of the opportunity to ride his bike while leaving other opportunities for action intact), in this version Black is like a conditionally activated puppet-master whose intervention is triggered by *any attempt*, by Jones, *to act in any way* contrary to Black's plan. Jones was not *actually* Black's puppet – he always made his own decisions, without intervention from Black – but he was always just a short step away from being Black's puppet. Still, Jones retained all his abilities, including the abilities that are relevant to moral responsibility – the ability to deliberate and to make decisions on the basis of moral reasons. Jones's abilities were highly fragile, during that year, and he was at constant risk of losing his ability to make his own decisions. But to be at risk of loss is not the same as having actually lost.

Black as Pure Counterfactual Intervener

In order to cash Frankfurt's promissory note, Black needs extra powers. Let's examine the case in which Black is a pure counterfactual intervener. There is a problem, though. Counterfactual intervention is possible only if it is possible to make reliable predictions of the actions of free and responsible agents. Deterministic agents are predictable (at least in principle), but the success of Frankfurt's thought experiment is not supposed to require the truth of determinism and it is controversial whether Black could make sufficiently reliable predictions of the free decisions of an indeterministic Jones. In my 2000 paper, I granted the defenders of Frankfurt that which makes their case the

strongest: that it is somehow possible to make ultra-reliable predictions about indeterministic events, including the decisions and actions of free and responsible agents. I argued that *even if* this is granted, Frankfurt's promissory note is not cashed. (And if it's not granted, Frankfurt's promissory note isn't cashed either.) I based my arguments on a story in which Black is a pure counterfactual intervener who somehow has the mysterious ability to reliably predict how an indeterministic coin will land, in the absence of intervention. Many people were confused by this story, asking questions like: "How can anyone know in advance what the outcome of an indeterministic process will be?" This is a problem for the defenders of Frankfurt, not me. If they want to use Frankfurt stories to undercut the traditional debate between compatibilists and incompatibilists, they need to show that pure counterfactual intervention is possible and that it works.

In what follows, I'm going to make things simpler for the defenders of Frankfurt by telling a story in which determinism is true so Black's ability to predict Jones's choices and decisions is not mysterious and counterfactual intervention is uncontroversially possible. If counterfactual intervention works at all, then it works in the deterministic case. I will argue that even under these conditions, counterfactual intervention is bogus. Insofar as Black is a pure counterfactual intervener, he does not alter *any* of the relevant modal facts about Jones. But, first, a note to forestall possible misunderstanding. My purpose here is to show that Frankfurt's thought experiment should never have persuaded any compatibilist to give up the belief that moral responsibility requires that a person could have done otherwise.[15] Some people (they used to be called 'incompatibilists' but these days they sometimes go by the name "semi-compatibilists" or "Frankfurt-style compatibilists") believe that it follows from the truth

of determinism that no one could ever have done otherwise. If this is your view, then you are not eligible for this version of Frankfurt's thought experiment. (And if you already believe that deterministic agents are morally responsible even though they could never have done otherwise, then you don't need the thought experiment.)

Suppose, then, that determinism is true but it is at least sometimes true that Jones could have done otherwise.[16] Black has the knowledge of a LaPlaceian predictor; he can predict everything that Jones will do before Jones is born, and his predictions are always right. Black's extensive and ultra-reliable knowledge permits Black to intervene well in advance to ensure that every detail of Jones' life goes according to plan. If we give Black enough power, Black can intervene early enough, and in subtle enough ways (eg. by altering the details of Jones' early childhood moral education) so that Jones' freedom and ability to do otherwise are left intact both at the actual world where Black does not intervene and the non-actual worlds where he does. But let's stick to the schema of Frankfurt's thought experiment and consider only those cases where Black would intervene by rendering Jones unable to do otherwise. Here's how he does it. In the morning Black makes predictions about what Jones will do (based on his brain-scanning device, plus information about the rest of Jones' environment). If he predicts that Jones will, in the absence of intervention, do something that is contrary to Black's plan, then Black stays on the job, ready to intervene at the appropriate time in a way that will prevent Jones from even beginning to do (decide, begin to decide, and so on) anything contrary to the plan. On the other hand, if Black predicts that Jones will comply perfectly with the plan, he retires for the day, and pays no further attention to Jones. This morning Black predicted that Jones will do all and only what Black wants him to do, so he retired

early. In the evening, when Jones decided, as Black knew he would, to go for a walk, Black was far away and fast asleep.

I hope it is intuitively obvious to you that in my story the facts about Black (on that particular day) made no difference to the facts about what Jones was able to do. If it isn't, there probably isn't much I can say to make you change your mind (again, see Vihvelin 2000), but let me walk you through the story anyway. It is early evening and Jones is at home. He has no plans yet for the evening. His friend Sally phones him and asks if he'd like to go for either a bike ride or a walk. Jones contemplates bike-riding but decides that he's not in the mood. He says "yes" to the walk, and off they go. On their way, they pass Jones's bike and Sally says "Are you sure….?" Jones again thinks about it for a moment, but again decides that he would rather walk. And so it goes, for the rest of the evening. Black is far away and sound asleep. The brain-scanning device has been turned off; there is no one and nothing monitoring Jones or prepared to stop him should he suddenly change his mind and decide to ride a bike after all (or do anything else contrary to Black's plan). Jones is as free as a deterministic agent ever is. Black has made no difference whatsoever to the facts about what Jones could have done.

Fischer would object that this story (like my story about the coin) is *not* like the Frankfurt stories told in the literature because Black is not on the scene monitoring the situation. Exactly. That was my point. I wanted a story in which it is vividly clear that Black is a *pure* counterfactual intervener and the only way to ensure this is to keep him off the scene. If we allow Black on the scene, it is too easy to think of Black as combining the powers of counterfactual and conditional intervention. (Cf. Fischer's

description of the babysitter who lurks outside the child's room, ready to stop him if he *tries* to come out [17] ; this is a description of a *conditional* intervener).

From what Fischer says about my coin story[18] , it appears that he would agree with me that on this particular day the facts about Black did not make it true that Jones was unable to do otherwise. But not all supporters of Frankfurt would agree. Here are some of the arguments I have heard:

1. We know in advance that Jones will do exactly what Black wants him to do.
2. One way or another Jones will end up acting according to Black's plan.
3. Given the facts about Black, Jones *must* act according to Black's plan; there are no possible worlds where these facts obtain and Jones does otherwise.

See Vihvelin 2000, for criticism of these and other fatalist arguments.

Why the Pure Counterfactual Intervener Cannot Deliver the Promised Goods

Back to my story. Here is Jones, in the evening, on his way out the door. He passes his bike, gleaming in the early evening sunshine and is, for just a moment, tempted to ride it. He remembers the joys of bike-riding, but he also remembers that it is more work than walking, especially with Sally, who likes to ride fast. He decides (as Black knew he would) to stick to his earlier decision to go for a walk.

Jones could have done otherwise. He could have accepted Sally's offer and gone for a ride on his bike. He had the ability; he knows how to ride a bike, and the relevant

parts of his brain and body were functioning correctly - no broken limbs, loss of muscle control, pathological fear of bike-riding, and so on. He also had the opportunity; his bike was right there, in good working order, and there was no Black hovering in the background ready to thwart his bike-riding efforts.

In this story of *pure counterfactual intervention,* Black has made no difference to the facts about Jones' abilities and opportunities. And Black has made no difference to the (relevant) counterfactual facts about Jones. If we wonder whether Jones had the opportunity, as well as the ability, to ride his bike, we ask whether Jones would have succeeded in riding his bike, if he had tried. [19] And the answer to that question is "yes". If Jones had tried to ride his bike (that evening, when Sally asked him, with the bike right in front of him), Black would not have been there to stop him, so he would have done so.

Granted, we could reason as follows: If Jones had tried to ride his bike, Black would (given his LaPlaceian knowledge of the earlier causes of Jones' decision) have predicted it. And if Black had predicted that Jones would try to ride his bike, he would (given his firm and unwavering intention that Jones *not* ride his bike), have taken steps to prevent it; he would be on the scene, ready to break Jones's bike or his legs or do whatever else it takes to stop him from riding the bike. And *if all this were true,* then Jones would not have succeeded in riding his bike.

This kind of "back-tracking" counterfactual reasoning has its place. We engage in it when we ask questions about counterfactual situations in which our primary interest is in facts about someone's *character and dispositions*; for instance, when we ask questions about what the past would have to have been like in order for it to be the case that a

cautious non-suicidal person jumps off a bridge, or in order for it to be the case that a proud person asks a favor of someone.[20] That's what's going on here, when we reason backwards from Jones' bike-riding attempt to how the past would have to have been in order for this to be consistent with the facts about Black's knowledge, and then forward again in a way that preserves the facts about Black's intentions and power. But this way of evaluating counterfactuals is illegitimate when we are evaluating *causal* counterfactuals; that is, in contexts in which our primary interest is in questions about the causal upshots of some nonactual event or state of affairs.[21]

Counterfactuals are notoriously vague. (If Caesar had been in Korea, would he have used the atom bomb or catapults?)[22] There are different ways of resolving the vagueness of counterfactuals; given a possible worlds semantics, this cashes out in terms of different similarity metrics we may use to pick out the relevant set of closest worlds. When we engage in the "back-tracking" counterfactual reasoning just described, we are using a similarity metric that says that the most important similarity respect is similarity with respect to the *character and dispositions of particular people.* There is nothing wrong with this, so long as we don't forget that this is what we are doing, and so long as we don't use this kind of reasoning to support a claim about a causal counterfactual.

Mary is standing on a bridge, without a parachute. Being a cautious, non-suicidal person, she does not jump, and her reasons for not jumping include her knowledge of the fact that *if she jumped, she would get hurt.* But there is a "back-tracking" argument that appears to deny the truth of this counterfactual causal fact:

(1) If Mary had jumped off the bridge, she would have first put on a parachute.

(2) If Mary had put on a parachute (before she jumped off the bridge), she would not have been hurt.

(3) Therefore, if Mary had jumped off the bridge, she would not have been hurt.

We can agree that (1) is true because Mary is so constituted as to always take precautions before she puts herself in danger, so the (relevant) closest worlds where Mary jumps off a bridge are all worlds where she has first arranged to be wearing a parachute. We can agree that (2) is true because Mary is a skilled parachutist, so the (relevant) closest worlds where Mary is wearing a parachute when she jumps are all worlds where Mary doesn't get hurt. And we can even agree that (3) is true provided we understand it the following way: The (relevant) closest worlds where Mary jumps are all worlds where she is wearing a parachute and retains the parachuting skills she in fact has.

But we would be making a mistake if we used this reasoning to conclude that (3) is true, given the way we ordinarily understand (3). For (3) is most naturally understood as a *causal counterfactual;* that is, as a claim about the causal upshots of Mary's jumping off the bridge, given the way things actually were, then and there. And it is false that if Mary had jumped, then and there, the causal upshot would have been a safe landing. For, as a matter of fact, Mary was *not* wearing a parachute. And if she had jumped, she would still not have been wearing a parachute. So if Mary had jumped, this would have caused her to get hurt.

The case of Jones is exactly parallel to that of Mary. Jones was standing right beside his bike when he considered, and turned down, Sally's proposal. If Jones had decided to ride his bike, his decision would have been causally efficacious, and he would

have done so. But there is a back-tracking" argument that appears to deny the truth of this counterfactual causal fact:

(1) If Jones had decided to ride his bike, Black would have known about it in advance.

(2) If Black had known in advance that Jones would decide to ride his bicycle, Black would have made it impossible for him to do so.

(3) Therefore, if Jones had decided to ride his bicycle, Black would have made it impossible for him to do so.

We can agree that (1) is true because Jones is so constituted as to give reliable prior signs of what he will decide to do and Black is watchful for those signs. We can also agree that (2) is true because Black is so constituted that he will intervene and rob Jones's decision of any efficacy. And we can even agree that (3) is true, provided that we understand it as saying that the (relevant) closest worlds where Jones decides to ride his bike are all worlds where Black makes a different prediction in the morning and is on the scene in the evening, ready to break Jones's legs or snatch his bike before Jones rides it. But we would be making a mistake if we used this reasoning to conclude that (3) is true, given the way we ordinarily understand (3), that is, as a *causal counterfactual.*

It is now widely accepted that causal counterfactuals are evaluated in the way that David Lewis has taught us: by a similarity metric that tells us to consider those worlds where the past is the same as the actual past until the occurrence of a small "divergence miracle" at or shortly before the time of the antecedent and which perfectly obey the laws of our world after the time of the antecedent. [23] This similarity metric yields the intuitively right result for the counterfactual about Jones. If Jones had decided to ride his

bike that evening, the facts about the morning would still have been what they actually were, and Black would still have predicted that Jones would not decide to ride his bike and so would still have retired for the day. So if Jones had decided to ride his bike that evening, he could and would have done so.

Fischer suggests that perhaps a "back-tracking" argument is legitimate if Black is a sufficiently reliable predictor ("the prior sign or triggering event would occur if and only if the agent were about to choose or do otherwise").[24] But this misses the point. We can grant that Black not only always gets it right but *must* always get it right; we already granted this when we made Black a LaPlaceian predictor. These facts about Black do not change the similarity metric for causal counterfactuals.

Fischer also suggests that a "back-tracking" argument is legitimate if Black is on the scene. [25] Again, this misses the point. To see this, consider the following case: Jones is depressed and suicidal and Black is following him around to make sure he doesn't hurt himself. Right now Jones is perched on the ledge of a high building. If he falls, he will plunge instantly to his death. Black has a safety net, but it takes time to set it up; if Black waits until Jones jumps it will be too late to save Jones. Luckily for Jones, Black is able to reliably predict Jones' actions shortly before they occur, and this gives him enough time to get the safety net in place. Today Black correctly predicts that Jones will not jump, so he doesn't bring out the safety net. Query: What would have happened if Jones had jumped? Answer: He would have plunged instantly to his death (since the safety net was *not* there). This case is exactly parallel to the two cases I just described. In each case, there is a "back-tracking" argument that may be used to support a counterfactual that appears to deny a causal counterfactual; in each case, there is nothing wrong with the

back-tracking argument, *in its place.* The only mistake is in using the "back-tracking" argument to support the causal counterfactual.

I conclude that Black, insofar as he is a *pure counterfactual intervener,* does not change *any* of the relevant modal facts about Jones; he neither robs Jones of any abilities, nor does he deprive him of opportunities, nor does he change any of the relevant causal counterfactuals true of Jones. This should not be surprising. After all, there is a big difference between the *existence* of a power and the *exercise* of the power. Black has the power to deprive Jones of all his abilities and opportunities, and if he were to exercise his power, these modal facts about Jones would be very different. But since Jones does exactly what Black wants him to do, Black never exercises his power. So these modal facts about Jones remain unchanged.

Counterfactual Logic, Counterfactuals, and Ability: Reply to Fischer

What I have said so far has been a clarification of what I said in my CJP 2000 and, in some places, an addition to it.[26] In this section, I will address Fischer's criticisms directly.

Fischer's criticisms all come down to claims about counterfactuals. He complains, first, that my story about Black as a pure counterfactual intervener is not like the Frankfurt stories told in the literature because Black is not on the scene so "the relevant counterfactuals – those allegedly parallel to the relevant counterfactuals in the Frankfurt stories – are not true." Second, he defends the "back-tracking" counterfactual argument I criticised, claiming that in the *right* kind of Frankfurt story, the argument is both valid and sound. Finally, he says that when Black is on the scene it's not clear that

any *argument* is needed because it is "intuitively obvious" that "if the relevant individual (Jones) were about to refrain, he would be rendered unable to refrain" and this makes it "intuitively obvious" that Jones is unable to do otherwise.

I have already addressed Fischer's first objection by explaining my reasons for removing Black from the scene. If we leave Black on the scene, it is too easy to forget that Black is supposed to be a *pure* counterfactual intervener and to endow him with the powers of a conditional intervener as well. More generally, I claim that Frankfurt stories are underdescribed, and thus badly designed, thought experiments because they fail to distinguish the two very different ways in which Black might be prepared to intervene. My purpose, in telling two separate "Frankfurt-style" stories, was to sharply distinguish the case in which Black has *only* the power of counterfactual intervention from the case in which he has *only* the power of conditional intervention. In order to achieve this, it was necessary to describe a case in which Black is an *ultra-pure* (off the scene) pure counterfactual intervener. By telling such a story, I hoped to make it crystal clear, to even the staunchest Frankfurt defender, that the *existence* of a pure counterfactual intervener makes no difference to the relevant modal facts. My thought experiment was successful. Fischer agrees that in my story it remains true that the coin could have come up tails and thus agrees that in the parallel story about Jones and an off-stage Black, Jones remains able to do otherwise. But Fischer nevertheless claims that the right kind of pure counterfactual intervener can deliver the goods that the conditional intervener could not deliver. What's needed, he maintains, is a pure counterfactual intervener who is on the scene, for only by *being there* can the pure counterfactual intervener make the "relevant counterfactuals" true.

We may wonder: What are these "relevant counterfactuals" and what, exactly, is the relevance of *any* counterfactual to the truth of claims about what a person can and cannot do? I will address these questions in a moment. But first I want to respond to Fischer's second criticism.

Fischer's second criticism is in response to the paragraph in my CJP 2000 that criticised a "back-tracking" argument in defense of the claim that the coin in my story *could not have landed tails*. Here is the argument:

COIN

1. If the coin were about to land tails, Black would have predicted this and intervened.
2. If Black had predicted this and intervened, the coin would have been forced to land heads.
3. So if the coin were about to land tails, it would be forced to land heads.

My criticism:

"But this objection relies on a form of counterfactual reasoning that is generally considered invalid: hypothetical syllogism."

Fischer agrees that the conclusion of **COIN** is false[27] and he agrees that hypothetical syllogism is not a valid inference form for counterfactuals. [28] Fischer's claims about **COIN** commit him to similar claims about the "back-tracking" argument I discussed in the previous section:

DECIDE

1. If Jones had decided to ride his bike, Black would have known about it in advance.
2. If Black had known in advance that Jones would decide to ride his bike, he would have made it impossible for Jones to do so.

3. Therefore, if Jones had decided to ride his bike, Black would have made it impossible for him to do so.

Fischer also agrees that the structurally parallel "back-tracking" argument sometimes invoked by Frankfurt defenders is "not formally valid".[29] His example:

ABOUT

1. If Jones were about to refrain, the triggering event would already have occurred.
2. 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.
3. Therefore, if Jones were about to refrain, he would be rendered unable to refrain.

Nevertheless, Fischer claims that the conclusion of **ABOUT** (and similar arguments used by Frankfurt defenders) "follows from the relevant premises together with other facts", and he claims that the "other facts" are supplied by Frankfurt stories. He concludes: "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."[30]

This is puzzling, because the background facts of **COIN** and **DECIDE** are not so different from the background facts of **ABOUT**. In all three cases, we are told that Black is a highly reliable predictor who knows in advance what a fair coin/free agent would (in the absence of intervention) do and we are also told that Black has a plan for the coin/Jones and the power and determination to enforce his plan, but only if he has to. If Fischer believes that the premises of **ABOUT,** together with the background facts**,** imply the "relevant conclusion", does he also believe that the premises of **COIN** and **DECIDE,** together with the background facts, imply the "relevant conclusion"?

Fischer never answers this question, but he does tell us that the first premise of **COIN** (and thus **DECIDE**) is false because Black is "far away and fast asleep", whereas the first premise of **ABOUT** is true because Black is wide awake and on the scene. I will examine this claim in a moment. But first, let's consider his argument for the validity of **ABOUT.**

Fischer has two different arguments for the claim that **ABOUT** is valid. His first argument is based on a diagnosis of several counter-examples to counterfactual hypothetical syllogism; he claims that these counter-examples have a "characteristic structure" which is not present in Frankfurt stories. What these arguments have in common, according to Fischer, is that the antecedent of the first premise is "more far-fetched" than the antecedent of the second premise, so the premises "send us" to different worlds (or sets of worlds). BBy contrast, where the argument is based on a Frankfurt story, the premises "send us" to a single world (or set of worlds). Fischer sums this up by saying that the "problematic feature" of arguments of the first kind is that we evaluate the premises by "world-hopping". [31]

But this diagnosis of what constitutes a counter-example to counterfactual hypothetical syllogism is not right. Consider this slight variation on Lewis's story about Otto. Waldo is so smitten with Anna that *he goes whereever she goes, even if Otto is also there.* As in Lewis's story, Otto is Waldo's successful rival for Anna's affections and Otto was locked up at the time of the party, but Anna almost went. Given these background facts, the following argument is *not* a counterexample to counterfactual hypothetical syllogism:

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

2. If Anna had gone, Waldo would have gone.
3. Therefore, if Otto had gone, Waldo would have gone.

But in this story, as in Lewis's original story, Otto's going to the party is more "far-fetched" than Anna's going and the premises "send us" to different worlds.

The difference between this story and Lewis's original story is that in my story an additional counterfactual is true:

3. If Anna and Otto had gone to the party, Waldo would have gone.

Whereas in Lewis's story it is true that:

2. If Anna had gone to the party, Waldo would have gone.

But false that:

3. If Anna and Otto had gone to the party, Waldo would have gone.

It is a well known[32] fact about counterfactuals that Antecedent Strengthening is not valid. That is, it is a counterfactual fallacy to reason:

A []→ C
Therefore, A & B []→ C

If Antecedent Strengthening were a valid pattern of counterfactual inference, there would be no difference between Lewis's Otto story and mine. It is the invalidity of Antecedent Strengthening that accounts for the invalidity of Hypothetical Syllogism. As Lewis puts it, "The fallacy of transitivity [and thus hypothetical syllogism] is a further generalization of the fallacy of strengthening the antecedent…Inference by transitivity would …justify inference by strengthening the antecedent; since we know that the latter is fallacious, so is the former." [33]

Fischer's second argument for the validity of **ABOUT** is explicitly based on Lewis's semantics and logic for counterfactuals. Following Lewis, Fischer says that a *sufficient* condition for the validity of a particular argument of the form

A []→ B, B []→ C, therefore A []→ C

is that we evaluate both premises in terms of the same set of worlds; that is, in the context in which the argument is asserted, the set of closest A-worlds is identical to the set of closest B-worlds.[34]

Lewis also makes a stronger claim. He says that the inference pattern

A []→ B, B[]→A, A []→ C, therefore B []→ C

is valid.[35] Fischer neither asserts nor denies this stronger claim, but he argues as follows:

ABOUT is valid because there is "just one world (or set of worlds) to which the premises send us". [36] **ABOUT** is sound because "the Frankfurt-story posits a single scenario in virtue of which the two premises are true". "[37]

No problem, so far. The problem is that Fischer denies that **COIN** and **DECIDE** are sound. But how can he say this? For it is *also true*, given the Frankfurt-style stories that I told about **COIN** and **DECIDE,** that we can understand the premises so that they "send us" to just one world or set of worlds – worlds where Black has the same powers, predictive abilities, and plan for the coin/Jones that he actually has. And it is *also true* that the two premises of **COIN** and **DECIDE** are true in virtue of this single "scenario" (world or set of worlds).

The real problem with "back-tracking" arguments like **ABOUT**, **COIN**, and **DECIDE** is not that these arguments are invalid or even that the premises *must* be read as false. The real problem is the one I explained in the previous section. When we accept the premises of these three arguments, we are moved to do so by the fact that we are

sympathetic and charitable listeners and we are inclined to understand a speaker's utterances in a way that makes them true. So we use a similarity metric that gives the most weight to similarity with respect to *the facts about Black* and, reasoning in this way, we agree that the premises – of all three arguments – are true. We may then be moved to think the conclusion is also true, and this temptation is especially strong in the case of **ABOUT.** But the "relevant conclusion" is a *causal counterfactual* – a claim about what would have happened, in the actual circumstances, if the coin had been about to land tails, if Jones had decided to ride his bike, if Jones had been about to refrain from doing X (where X is the action Black wants him to do). And causal counterfactuals are evaluated by a different similarity metric, one that doesn't give any special significance to the facts about any one person, not even the powerful Black. This similarity metric picks out a different set of worlds – worlds at which the past is the same as the actual past and Black made the same prediction that he actually made (and thus lacked the foreknowledge that he actually had). And, from the standpoint of *this* set of worlds, the conclusion of all three arguments – **COIN, DECIDE,** and **ABOUT** – is false. If the coin had been about to land tails, Black would still have predicted that it was going to land heads, so he wouldn't have forced it to land heads . If Jones had decided to ride his bike, Black would still have predicted that he wouldn't, so he wouldn't have stopped him. If Jones had been about to refrain, Black would still have predicted that Jones would do what he wanted him to do, so he wouldn't have forced him to act, rendering him unable to do otherwise.

Fischer's third and final objection is that if we describe the *right kind of case*, then we don't need a "back-tracking" argument to support the causal counterfactual that is the

desired conclusion of **ABOUT** because it is "intuitively obvious" that "if the relevant individual (Jones) were about to refrain, he would be rendered unable to refrain" and this makes it "intuitively obvious" that Jones is unable to do otherwise.

I do not doubt that Fischer correctly reports his intuitions, but after almost forty years of debate, it is clear that this particular question cannot be settled by appeal to intuition. If we all shared Fischer's intuitions, we would by now have all agreed that Frankfurt stories succeed in showing that **PAP** is false. So we need to move beyond intuitions, and that is what I have done, in this paper.

But let's take a closer look, anyway. Fischer reports two intuitions. He has the intuition that it is true that:

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

And he has the intuition that, because of **VERGE**, it is true that:

UNABLE: Jones is unable to do otherwise.

For the sake of argument, let's grant Fischer that, in the right kind of story, **VERGE** is true. The question is whether **VERGE** grounds, licenses, or in any way supports **UNABLE.**

There is an intermediate step that Fischer omits. If Jones is unable to do otherwise, it is presumably because he is unable to refrain from performing the action that Black wants him to perform. So if **VERGE** supports **UNABLE** it is because **VERGE** supports:

UNABLE (refrain): Jones is unable to refrain.

To refrain from performing an action is to act intentionally, so if Jones is unable to refrain from doing what Black wants, there is some action or way of acting -- call it R – such that Jones is unable to do R.

Is there any reason to think that the truth of **VERGE** grounds, licenses, or in any way supports **UNABLE (refrain)**?

Note that **VERGE** and **UNABLE (refrain)** say different things. **VERGE** says that *if things were different from the way they actually were* – if Jones were on the pre-action verge of doing something (R) that he didn't actually do and was never on the verge of doing – then Jones *would be unable* to do that thing. **UNABLE** says that Jones is *in fact unable* to do that thing.

But we do not ordinarily reason in this way, from a counterfactual about someone's inability to do something, to the conclusion that they are in fact unable to do that thing. For instance, we can agree that:

BROKEN LEG: If my leg were broken, I would be unable to walk.

But since my leg *isn't* broken, we don't conclude that I am *in fact unable* to walk.

And we can agree that it is true that:

BROKEN BRAIN: If my brain were broken, I would be unable to think.

But since my brain isn't broken, we don't conclude that I am *in fact unable* to think.

We might also agree that it is true that:

HYPNOTISED: If I were hypnotised, I would be unable to refrain from acting on the hypnotist's commands.

But since I am not hypnotized, we don't conclude that I am *in fact unable* to refrain from acting on the hypnotist's commands.

Why should we treat **VERGE** differently from **BROKEN LEG, BROKEN BRAIN**, **or HYPNOTISED?**

You might be tempted by the thought that the antecedents of **BROKEN LEG, BROKEN BRAIN**, **and HYPNOTISED** describe situations remote from actuality, whereas the antecedent of **VERGE** is true at a nearby world. But while this might be a factor that affects our intuitions, it's not a good reason. We can describe situations where the antecedents of **BROKEN LEG, BROKEN BRAIN, AND HYPNOTISED** are true at nearby worlds. I'm on a battlefield or driving a fast car. I'm a hypnotizable subject, and I just volunteered, but the hypnotist didn't pick me. In these situations, it's true that if things were just a little bit different from the way they actually were, I *would be caused to be unable* to walk, or to think, or to refrain from acting on the hypnotist's commands. And it's true that I am *at risk* of being rendered unable to do those things. But it's *not* true that I am *in fact unable* to do any of these things.

Is there any other argument open to Fischer? I can think of only one. Fischer might claim that Jones is able to refrain only if it is true that *if he had wanted to refrain, he would have done so*, but this counterfactual is false, because Jones' wanting to refrain is the mental event that triggers Black's intervention. That is, Fischer might argue that **VERGE** supports **UNABLE (refrain)** because there is a conceptual connection between a person's being able to do something and the fact that if he had wanted to do it, he would have done that thing. But this is not a good move for Fischer, for two reasons. First, almost everyone now agrees that it's a mistake to think that there is a conceptual connection between claims about what someone is able to do and counterfactual conditionals of this sort. [38] Second, and more important, Frankfurt's aim was to

undermine the traditional debate between compatibilists and incompatibilists by telling a story designed to convince us that **PAP** is false regardless of what we mean by locutions like "could have done otherwise" and "is able to do otherwise." If the Frankfurt defender needs to appeal to claims about the meaning of "could have done otherwise" in order to defend the claim that Frankfurt stories succeed as counterexamples to **PAP,** then the Frankfurt strategy for defending compatibilism fails.

Conclusion

In my CJP 2000, I stressed the difference between freedom of action and freedom of will, arguing that a conditional intervener removes the former but not the latter, so Frankfurt defenders need to show that a *counterfactual intervener* succeeds in removing Jones' *freedom of will.* But, I argued, insofar as Black is a pure counterfactual intervener, he doesn't succeed in making *any* difference to the facts about what Jones is free to do, either with his mind or with his body. I now think that Frankfurt stories go even more radically wrong. Insofar as Black is a conditional intervener, he does succeed in removing what we ordinarily think of as freedom of action. But that's because we think of this as *ability plus opportunity.* Qua conditional intervener, Black deprives Jones of the *opportunity* to succeed in doing anything other than the things Black wants him to do, but Black leaves all of Jones' *abilities* intact. Jones' abilities include the ability to do things with his body (to ride his bike, to speak English, and so on) as well as the ability to do things with his mind (to rehearse a speech, solve a math problem, and so on). And they include the abilities that have traditionally been thought to be necessary for free will – the ability to choose and to act on the basis of one's own reasons and reasoning; the

ability to be the agent-cause of one's actions, and so on, for whatever other ability you think is relevant.

What about the case where Black is a counterfactual intervener? The same point stands. Since Black never lays a finger on Jones, the *most* that Black, qua counterfactual intervener, can do is to deprive Jones of further opportunities. I have argued that we have no reason to believe that Black succeeds in doing even this much; insofar as Black is *only* a counterfactual intervener, he neither robs Jones of any abilities nor does he deprive him of any opportunities. But even if you are not convinced, you must agree that Black doesn't rob Jones of his *ability* to deliberate, to decide, to refrain, or to perform any action whatsoever.

Once we've grasped this fact, the lesson of Frankfurt stories turns out to be surprisingly simple. To remove someone's abilities, you must mess with their brain or body. To have the ability to do something is to have "what it takes" to do that thing, and whether this is true or not depends on what *you* are like, and not on facts about your circumstances. Facts about your circumstances, including the unfortunate fact that Black is hovering in the background, affect what counterfactuals are true of you, and may affect your opportunities, but they do not affect what abilities you have.

The conclusion we must draw is that Frankfurt stories fail. Jones loses some of the freedom we ordinarily have – and value – the opportunity to *successfully act* in alternative ways. This shows that moral responsibility requires *less* than all the freedom that most of us want and value. But he loses none of the freedom traditionally thought essential to free will; he retains the ability to make his own decisions, including decisions other than the decisions that Black wants him to make.

Frankfurt's article was written in 1969, at a time when criticisms of the "Conditional Analysis" of "could have done otherwise" had convinced many compatibilists to give up the attempt to defend the claim that our commonsense of talk of 'can' , "has the ability', and 'is able to' can be analysed in terms of counterfactual conditionals. (Ironically, this project was abandoned at the very moment when work on possible worlds semantics and the Lewis/Stalnaker semantics for counterfactuals was about to provide philosophers with the logical tools required to make progress on the modal and metaphysical issues needed to properly evaluate and respond to these criticisms.) Frankfurt's thought experiment convinced a generation of philosophers that these difficult metaphysical questions could be bypassed. But now, almost forty years later, Frankfurt's heirs are still attempting to cash Frankfurt's promissory note and in their attempts to do so are engaged in arguments about counterfactuals and metaphysics of the very sort that Frankfurt had hoped to avoid. If we are going to do metaphysics anyway, and especially if we are going to engage in debate about the evaluation of counterfactuals and the relevance of counterfactuals to claims about the abilities of agents, then I suggest that we stop talking about Frankfurt's thought experiment and start talking again about the traditional metaphysical questions: Does determinism deprive us of the abilities required for free will and moral responsibility? How should we understand these abilities? What is the relation between abilities and counterfactuals? Between abilities and dispositions?[39]

Kadri Vihvelin

University of Southern California

[1] See, for instance, "The Modal Argument for Incompatibilism", *Philosophical Studies* 53 (1988), 227-244; "Freedom, Necessity, and Laws of Nature as Relations between Universals", *Australasian Journal of Philosophy* 68 (1990), 371-381; "Freedom, Causation, and Counterfactuals", *Philosophical Studies* 64 (1991), 161-184; "Stop Me Before I Kill Again", *Philosophical Studies* 75 (1994), 115-148; "Free Will Demystified: A Dispositional Account", *Philosophical Topics* 32, *Agency*, (2004), 427-450; and "Compatibilism, Incompatibilism, and Impossibilism", in *Contemporary Debates in Metaphysics*, edited by Theodore Sider, John Hawthorne, and Dean Zimmerman, Oxford: Blackwell, 2008.

[2] "Freedom, Foreknowledge, and the Principle of Alternate Possibilities", *Canadian Journal of Philosophy* 30 (2000), 1-24.

[3] In this paper, I will be using 'was able to do otherwise', 'was free to do otherwise', and 'could have done otherwise' interchangeably. I follow Frankfurt in making no assumptions about what we mean, or should mean, when we use these locutions. (But see notes 6, 7 and 10.)

[4] A "back-tracking" argument is one with the following pattern of reasoning: " If the present were different in way D, then the past would have been different in way E; if the past were different in way E, then the future would be different in way F; therefore if the present were different in way D, the future would be different in way F". Although the argument I criticised is a "back-tracking" argument, I did not use that term in my CJP 2000. I will say more of this later.

[5] "Freedom, Foreknowledge, and Frankfurt: A Reply to Vihvelin"

[6] I understand the ability/opportunity distinction in the way that it's understood in ordinary English. When we say that someone has the ability but not the opportunity to do something, we are invoking a contrast between a *person's* contribution to the facts that enable her to do something and the contribution made by the person's *environment or situation.* More precisely, the contrast is between the contribution made by the *intrinsic* properties of a person and the contribution made by the *extrinsic* or relational properties of a person. To have the ability to do something, is, roughly, to have "what it takes" to do that thing (eg bike-riding skills and unbroken limbs and properly working brain and …); to also have the opportunity is to be in a situation which provides you with what you need in order to exercise your ability (eg. a bike and….) and in which nothing extrinsic to you would prevent you from exercising your ability. (eg. no prison walls or chains or Black lurking in the background). In saying that the ability/opportunity distinction is embedded in commonsense, I am not claiming that commonsense can settle the question of whether we *really* have the abilities and opportunities we think we have or whether these abilities and opportunities are compatible with determinism.

[7] "Ability" is sometimes used in a weaker way, to mean only that someone has the know-how, skills, or competence, required to do something. (In the literature, this is sometimes called a "general ability".) Given this weaker sense of 'ability', a person with a broken leg retains the ability to ride a bicycle. It should be fairly obvious that Black doesn't rob Jones of his bike-riding ability in this weak sense. What is less obvious, but also true, is that Black doesn't rob Jones of his ability to ride his bicycle in

the stronger sense that we have in mind when we say, on a particular occasion, that a person has the ability to ride a bicycle: in addition to having bike-riding skills and competence, it is also true that the person has "what it takes" to exercise those skills, then and there. (See note 6.)

[8] The story I just told is not intended to be a Frankfurt story. A Frankfurt story is a story where i) we are supposed to agree that a person *could not have done anything* other than what he actually did; and ii) we are supposed to have the intuition that he is nevertheless morally responsible for *something that he did.* My story doesn't satisfy the first criteria. Black's interest in Jones is limited to his desire that Jones not ride his bike that year, so he leaves Jones as free as he ever was to perform other actions, including mental acts like deliberating about bike-riding, deciding to ride his bike, and so on. My aim in telling this story is not to make any point about moral responsibility, but, rather, to make the following point about counterfactuals and ability: We are not entitled to infer, from the fact that *someone would prevent you from doing X*, that you *in fact lack the ability to do X.* It may be that you retain the ability, and only lack the opportunity. And, for all that we've seen, it may be that you retain both ability and opportunity. I will say more of this latter possibility later.

[9] Harry Frankfurt, "Alternate Possibilities and Responsibility", *Journal of Philosophy* 66 (1969).

[10] I say "unfortunately" because Frankfurt's choice of name has led some philosophers to suppose that the commonsense platitude that PAP is supposed to capture is to be understood in terms of a person's "alternate possibilities". But of course alternate possibilities need not be actions, let alone actions that the person has either the ability or the opportunity to do. Recognition of this has led to talk of "robust" alternatives, or "genuine access" to alternatives, but these locutions are metaphorical and unhelpful. We can avoid this kind of confusion if we understand PAP as saying that a person is morally responsible for what he has done only if the person could have intentionally *acted* in some alternative way, and if we understand "could" as "was able to", leaving it open whether this should be understood as 'had the ability' or 'had the ability and also the opportunity'. (See notes 6 and 7.)

[11] For a classic statement and defense of the so-called "Conditional Analysis" of 'could have done otherwise', see G.E. Moore, "Free Will" in his *Ethics*, Oxford: Oxford University Press, 1912. For critique of the Conditional Analysis, and an alternative compatibilist account, see Keith Lehrer, "'Can' in Theory and Practice: A Possible Worlds Analysis", in *Action Theory*, ed. M. Brand and D. Walton, Dordrecht Reidel, 1976. For more references, see note 16.

[12] Ibid, p.8.

[13] "Freedom, Foreknowledge, and the Principle of Alternate Possibilities", this Journal, *ibid*.

[14] If it makes sense to talk of the opportunity to *begin or try to act*, then Jones retains the opportunity to begin or try to act contrary to Black's plan.

[15] I don't mean to imply that I take back what I said in my CJP 2000 about free indeterministic agents. It is controversial whether counterfactual intervention is possible where Jones is a free indeterministic agent, but I continue to insist that *if it is possible,* it does not work as advertised.

[16] Due to the enormous influence of Frankfurt, this view is now a minority view, even among those philosophers who call themselves compatibilists. But see David Lewis, "Are We Free to Break the Laws?", *Theoria* (1981) 47, 113-121; Michael Smith, 1997, "A Theory of Freedom and Responsibility", in G. Cullity, ed., *Ethics and Practical Reason*, New York, Clarendon Press; Michael Smith, 2004, "Rational Capacities", in *Ethics and the A Priori: Selected Essays on Moral Psychology and Meta-Ethics*, New York: Cambridge University Press; Joseph Keim Campbell, Compatibilist Alternatives," *Canadian Journal of Philosophy* 35 (2005): 387–406; Kadri Vihvelin, "Free Will Demystified: A Dispositional Account", *ibid*, and Michael Fara, "Masked Abilities and Compatibilism", *Mind*, forthcoming.

[17] Fischer, p.12.

[18] Fischer, pp. 8-11.

[19] This is not intended as either a definition or a necessary condition of opportunity but as defeasible evidence of opportunity. And it states a condition stronger than we ordinarily think is required for opportunity; we would ordinarily grant that Jones has the opportunity, as well as the ability, to ride his bike if we believe that if he tried to ride it, he *might* succeed, or *would have a reasonably good chance* of succeeding (or something like that).

[20] David Lewis, "Counterfactual Dependence and Time's Arrow", *Nous* 13 (1979), 455-476.

[21] Lewis calls this "the standard resolution" (Lewis, *ibid*), but that's because he wants to use counterfactuals to analyse causation, and because he ambitiously hopes to do this without making any assumptions about causation; in particular, he does not want to assume that the standard resolution for evaluating counterfactuals is the one appropriate for causal counterfactuals. If we reject this ambitious aim, then we can arrive at Lewis's widely accepted similarity metric for "standard" counterfactuals by a different route: we can say that our default or standard way of evaluating counterfactuals is the way that is appropriate for causal counterfactuals, and Lewis's similarity metric is the right one for causal counterfactuals. See note 23.

[22] This is Lewis's example. Lewis says: "It is right to say either, though not to say both together. Each is true according to a resolution of vagueness appropriate to some contexts." (Lewis, *ibid*, p. 34)

[23] John Collins, Ned Hall, and L.A.Paul, "Counterfactuals and Causation: History, Problems, and Prospects" in *Causation and Counterfactuals*", edited by Collins, Hall, and Paul, Cambridge, Mass.: MIT Press, 2004. See also David Lewis, *Counterfactuals, Harvard:* Harvard University Press, *1973; Lewis,* "Causation", *Journal of Philosophy* (1973) 70, 556-567; Lewis, "Counterfactual Dependence and Time's Arrow"; *ibid,* Lewis, "Are We Free to Break the Laws?", *ibid*, and many other papers.

[24] Fischer, p. 18.

[25] Fischer, p. 11.

[26] The chief additions are my use of the ability/opportunity distinction and my discussion of the uses and abuses of "back-tracking" counterfactual arguments.

[27] Fischer, p. 11.

[28] Fischer, p. 13.

[29] Fischer, p. 13.

[30] Fischer, p. 13.

[31] Fischer, pp. 14-15.

[32] Jonathan Bennett, *A Philosophical Guide to Conditionals*, Oxford: Clarendon Press, 2003, pp. 159-163 and pp.172-176. Bennett says that the invalidity of Antecedent Strengthening is "the securest thing we know about subjunctive conditionals". (p.169).

[33] *Counterfactuals*, 1973, *ibid*, p. 32. See also pp. 10-18.

[34] Fischer, p.15.

[35] Lewis 1973, pp. 33-34. Whether or not Lewis is right about this is a disputed point in the logic of counterfactuals. See Robert Stalnaker, *Inquiry*, Cambridge, Mass.: MIT Press, 1984, pp. 130-132, for a discussion of a putative counterexample due to Pavel Tichy. And see Jonathan Bennett, *ibid*, pp. 298-301, for a discussion of the semantic difference between theories of counterfactuals which accept the validity of this argument pattern and those which reject it.

[36] Fischer, p.15.

[37] Fischer, p. 17.

[38] For a survey and discussion of criticisms of the Conditional Analysis, see Bernard Berofsky, "Ifs, Cans, and Free Will: The Issues", in *The Oxford Handbook of Free Will*, Robert Kane, ed., Oxford: Oxford University Press, 2002.

[39] I am grateful to Joe Campbell, John Carroll, Charles Hermes, Terrance Tomkow, four anonymous CJP referees, and the commentators at the first Online Philosophy Conference for helpful comments on earlier versions of this paper.